Molecular Dynamics Simulations of Protein-Ligand Complex Systems and its Applications in Drug Design

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von
Min Xu
aus China

Promotionskomitee
Prof. Dr. Amedeo Caflisch (Vorsitz und Leitung der Dissertation)
Prof. Dr. Peter Hamm
Prof. Dr. Sereina Riniker

Zürich 2016

Summary

Understanding the patterns of protein-ligand interactions is of great importance in biological and medical research. These processes are intrinsically dynamic, usually involving multiple states and forming complex transition networks, which poses a big challenge for investigations. In this thesis, we employ molecular dynamics (MD) simulations to extend our knowledge in this field and assist in practice for drug discovery as well. Three complex systems have been explored here, namely FKBP/DMSO, PDZ domain/peptide and CRBBP bromodomain/small-molecule inhibitors. The corresponding proteins have diverse functions, ranging from rotamase, signal transduction to histone code reader, for all of which ligand recognition is essential.

It is critical to know the states of the protein when studying its dynamic properties for binding. Since DMSO is easy to dissociate from FKBP, only by MD simulations, we can equilibrate the protein in its complex form. The unbinding of DMSO causes a sizeable effect on the protein structure, which in turn greatly changes the binding behavior. We show that it takes a few hundred nanoseconds for the protein to reach equilibrium upon this perturbation and it is much slower than the unbinding itself. The implied capability of structural memory is worth consideration when we study the kinetics profiles for drugs.

Peptide-protein interactions could be extremely complicated because peptides usually possess much more degrees of freedom. With the help of powerful computer hardware and analytic tools, we are now able to obtain comprehensive pictures for this kind of systems. In particular, we find that the peptide in our study is guided by nonspecific electrostatic interactions during the initial association with the PDZ domain. Formation of encounter complex is identified along the pathway, which features completely different interactions from the final stereospecific complex. We think that this interaction pattern can be a general mechanism for many protein-peptide systems.

Finally, MD simulations provide valuable information in our virtual screening campaign for CREBBP bromodomain inhibitors. Short simulations are useful for compound selection and long simulations help to reveal binding modes of hits. Two different scaffolds of ligands are first discovered by high-throughput docking. Using MD-guided optimization, we can easily modify the compounds *in silico* and test the effects. In this way, we acquire a series of derivatives with nanomolar potencies, excellent ligand efficiency and selectivity.

Contents

Chapter 1

Introduction

People have realized for a long time that the functioning mechanism at the molecular level is the key to understand how things are going on inside a living organism. Biological macromolecules, such as proteins, nucleic acids, lipids and polysaccharose, form a vast interaction network and control every aspects of life entities. Molecular modelling, especially molecular dynamics (MD) simulation, is designed to investigate in this area. It helps us interpret experimental measurements, design novel proteins and drugs and predict biomolecular behaviors. MD simulations provide us with the extremely high-resolution view of the dynamic processes related to the biological macromolecules, which might be beyond the capability of any current experimental approaches. Besides, numerous *in silico* studies show good agreements with the results obtained *in vivo* or *in vitro* and thus prove the feasibility and effectivity of the computational methods. Ever since the very first application in protein BPTI [1], the impact of molecular simulation in the research of life science has become unneglectable and profound. New computer technology and data analytical approaches have continuously contributed to the development in this field, which makes it even more appealing and exciting.

1.1 What is molecular dynamics simulation?

MD simulation is about simulating the movements of the atoms and molecules following certain physical rules, for example, classical Newton's law of motion. To run an ordinary MD simulation, all we need is the initial coordinates of all the atoms in our system and a force field that defines the potential energy of the system. Knowing the potential energy, we can derive the forces on each atoms and thereby predict the moving trajectories by solving the Newton's equation as a function of time:

$$m_i \frac{\partial^2}{\partial t^2} \boldsymbol{r_i} = m_i \boldsymbol{a_i} = \boldsymbol{F_i} = -\frac{\partial}{\partial \boldsymbol{r_i}} E(\boldsymbol{r}) \tag{1.1}$$

where m_i is the mass of atom i, $\mathbf{r}_i$ is the coordinate of atom i, $\mathbf{a}_i$ is the acceleration of the atom i, $\mathbf{F}_i$ is the force acting on atom i and $E(\mathbf{r})$ is the potential energy of the system.

It might seem that this formula is too simple to study the biological systems. However, people have put so much effort in developing the force field as to make sure that it can produce an sufficiently good approximation [2]. Important biophysical phenomenon, such as solution-solvent diffusion, protein folding/unfolding and ligand binding, can be reproduced in the simulations just like those in the experiments. Furthermore, the most commonly used force field like AMBER [3], CHARMM [4], GROMOS [5] and OPLS [6] have been optimized and validated in many cases over the years.

Along with the improvement of the force field, the state-of-the-art technology in computer science also makes MD simulation more powerful in applications. Only couple of years ago, it was not an easy task to simulate a system for a few microseconds. However, most interesting biological events happen in the time scale range from hundreds of microseconds to seconds. Therefore, computational power used to be the biggest bottleneck. Today, normal CPUs are so fast that it becomes a routine job to conduct a microsecond-long simulation. Moreover, cloud computing [7], GPU acceleration [8] and Anton, a supercomputer designed specifically for MD simulations [9], they have pushed simulations to millisecond-long range and provided us with more chances to study the real sophisticated processes in atomistic details [10, 11].

With the rapid growing of the simulation data size and complexity, comprehensive analytic tools are developed to facilitate post-processing steps. A scalable algorithm was created in house to identify the kinetically distinct basins [12]. It extracts the kinetic information by calculating the distances among the snapshots and re-organizes them in a way such that kinetically close snapshots can be together to form states. Furthermore, we can add the general analytical measurements like RMSD and dihedral angle changes to aid annotation of the trajectories. Combining everything together, we designed a figure called SAPPHIRE plots to visualize the final result [13]. In my study of the CREBBP bromodomain inhibitors (Chapter 4), I used this approach to find out three different binding modes of the ligand, which has provided us with the valuable clue in optimization for a better binding affinity.

1.2 Application in the protein-ligand system

Understanding protein-ligand complex system is the fundamental part in the modern drug discovery and design projects [14, 15], so it is no wonder that MD simulations have been broadly used in the related researches.

People used to picture the protein-ligand system as “lock and key”. However, this model has severe limitation in practice because we now know that the binding of drugs or other ligands to protein is never static [16, 17]. Conformational selection and induced fit are two concepts that are widely accepted currently (Figure 1.1). The former believes that the bound conformations of receptor pre-exist and the interaction with the ligands would gradually shift the ensemble equilibria towards these states [18-22]. However, if the bound conformations cannot be found in any apo-structures, we think that induced fit is the dominant effect [20, 22, 23]. Despite of the differences between these two scenarios, it is more important that they both agree on one fact: binding processes should involve multiple states and transitions.

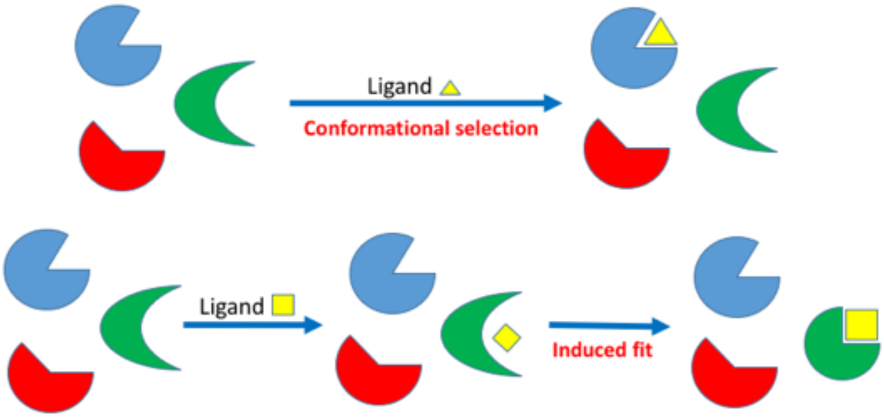


Figure 1.1 Illustration of conformational selection and induced fit [24].

It is critical for us to disclose all the intermediate states. One reason is that different protein conformations usually demonstrate distinct behaviors when interacting with the ligands. For example, in the research of the serine protease Trypsin and its competitive inhibitor Benzamidine, Ignasi *et al.* found out that different metastable structures of Trypsin possess different binding affinities and kinetics for binding [25]. This result is verified by Nuria *et al.* in their recent paper [26]. Their work also indicates that the protein undergoes a significant conformational change upon ligand binding, which cannot be described by a simple two-state model. The second reason is that kinetic property of the ligands is a key factor in drug development. Fast association rate raises the chance for binding the transient conformations, which might be important for selectivity. Slow dissociation increases the duration of drugs and fast dissociation is useful for the safety issues [27, 28]. Hence, by uncovering the elusive binding pathways, we can make smarter decisions in the rational design of drug by targeting some particular protein states.

Nevertheless, it is often the case that the interconversion among different states is either ultrafast or results from rather small conformational changes of proteins. It is certainly a difficult mission to disclose all the states and experimental techniques are not suitable for it. As already shown in the case of Trypsin, researchers often seek help from MD simulation to achieve their goal.

Due to the limited computational resources before, biased MD simulation like steered MD simulations was the only option for many systems [29-31]. These methods accelerate the association or dissociation process by modifying the potential energy and exerting external force on the ligand. Researchers still rely on them in many cases because they are very useful in discriminating binders from nonbinders and identifying the key residues. However, the weakness is also obvious: we need the prior knowledge about the correct binding poses; setting up the simulation system might be rather tricky; more importantly, the outcomes might be ambiguous since the free-energy landscape has been altered.

Unbiased simulation is more favorable but it was only applicable to ligands with small sizes since their binding/unbinding rates are usually fast and the conformational space to be explored is relatively small. For instance, our group did a research on the FKBP-DMSO complex and acquired the similar binding/unbinding kinetics profile as the experiments [32]. With the computational power today, we can use the same system to study a more subtle subject: protein and water relaxation (Chapter 2). To get a statistically convincing result, we finished more than 100 simulations starting from independent bound states until restoring the equilibrium of the unbound states. Moreover, we constrained the backbone of the protein in the additional simulations and discovered a difference response in water density change. This kind of research is simply not possible without MD simulations.

With the help of the advanced technology, the unbiased MD simulations start to show its value in more and more important studies. In 2011, the D.E.Shaw research group published their study on investigating the binding processes of alprenolol to β$_2$-adrenergic receptor (β$_2$AR, a G-protein-coupled receptors) using all-atom unbiased simulations [33]. This is a tremendously large system consisting of a 500-residue protein with hundreds of thousands of water molecules and lipids. In spite of the complexity, they succeeded in reproducing the crystallographic binding pose of alprenolol. Further analysis reveals a clear binding pathway with three stages: entering the extracellular vestibule, residing in the vestibule and inserting into the binding pocket. The free energy barriers suggest that the dehydration of the ligand and protein upon binding is as important as the final entry into the binding site through a narrow gateway. This kind of observation is crucial not only in the theoretical sense but also in practice for optimizing the drugs. Similar work on the same target was also performed using the cloud computing by Diwakar *et al.* [34]. By building the transition pathway of the binding process, they selected a few relevant conformations of the receptor and tested them with the conventional docking procedure. The outcome was very promising for that the selected structures significantly outperformed the crystal structures and random MD snapshots.

Aside from all the successful stories, there are still two obstacles keeping MD simulations from broader application. Firstly, to find a drug candidate, we typically have to investigate thousands, if not millions, of small molecules. This amount of work is apparently unrealistic for any kinds of MD simulations so far. The second one is a bigger headache: we do not have an accurate force field for small molecules as the biomolecules have. Small organic molecules, despite of their sizes, mostly occupy much larger chemical space [35]. Although people commonly use force fields like general AMBER force field (GAFF) [36] and CHARMM general force field (CGENFF) [37, 38] to do the parametrization, the performance of these force field is only goodish on average. We are very likely to get incorrect partial charges or topology for small molecules with complex motifs, which would make simulations meaningless.

1.3 Application in the protein-peptide system

Peptide-protein interactions play a prominent role in numerous biological activities, involving signal transduction, regulation of gene expression and activation of enzymes. Various vital pathways are operating and/or collaborating by using peptides to deliver the messages. It is no wonder that protein-peptide interactions have been exploited in various biotechnological and pharmaceutical applications, such as peptide-based fluorescent biosensors [39], biomarkers [40], antibody generation [41] and peptide drugs [42]. Therefore, revealing the mechanism behind protein-peptide interactions has great significance.

Studies on different protein-peptide complexes demonstrate that electrostatic interactions are the driving force for peptide recognition. Moreover, before the system forms the stereospecific complex, it constitute ensembles of structures called “encounter complex” [43, 44]. However, alternative binding modes are also suggested by some protein modules like PDZ domains [45-47].

PDZ domains are protein binding domains that are vital to signaling, scaffolding and trafficking in neural networks [48]. People used to think that PDZ domains can only recognize the carboxy-terminus of the partner protein, in which the charged C-terminal residue is the key for binding [49]. Finding out that PDZ domains can also binds some specific internal peptide motifs, people realize the possibility of an utterly different binding mechanism. Experiments on salt dependency and mutations imply that the steric and hydrogen bonding complementarity is the more primary factor in this kind of binding events [50].

To deepen the knowledge of PDZ-peptide interaction pathways, we performed unbiased MD simulations to investigate the association and dissociation of PDZ2 with a C-terminal hexapeptide (Chapter 3). Our simulations reproduce the exact binding pose in the crystal structure. Using the SAPPHIRE plot and other analytic methods, we construct the complete kinetic and free energy profile. It shows that the recognition is triggered by electrostatic steering. However, when the encounter complexes form, nonnative salt bridges become the main force for stabilization. Our conclusion connects the previous theories and demonstrates the interplay between different forces. In addition, it might be interesting to study the association kinetics of the internal peptides in the future and do a comparison with the current results.

Our study is just one single case among the extensive utilization of MD simulations in research of protein-peptide complexes. Like what has been shown for the small molecules, MD simulations has the advantage over experiments by disclosing the very details of the binding process. Besides, most peptides are composed of native amino acids, so we are freed from the pitfalls of force fields. Nevertheless, the biggest challenge remains in the fact that peptides can possess up to fifty residues, which is equivalent to a huge number of degrees of freedom. Therefore, the energy landscape may become so complicated that the simulations might easily get trapped in the local basins. In order to adequately sample the system, we need either more computational time or efficient sampling methods.

Replica exchange is one of the enhanced sampling approaches for such cases [51]. It helps the system to cross the barriers by running multiple replicas in parallel under slightly different temperatures and exchanging them periodically. The highest temperature makes the system escape the basins easily and the lowest temperature generates the samples we want. Our group developed another simulation protocol called PIGS last year. It avoids the redundant sampling in the same state by regularly reseeding the simulations from the sampling frontier [52]. Tested on some simple systems, it shows the capability to sample more states and have a better overlap with the canonical sampling than replica exchange. Therefore, studying on protein-peptide systems would be a good test for PIGS to prove its potential.

Bibliography

1. McCammon, J.A., B.R. Gelin, M. Karplus, Dynamics of Folded Proteins, Nature **267** (1977) 585-590.
2. Karplus, M., G.A. Petsko, Molecular Dynamics Simulations in Biology, Nature **347** (1990) 631-639.
3. Cornell, W.D., P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, Journal of the American Chemical Society **117** (1995) 5179-5197.
4. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, The Journal of Physical Chemistry B **102** (1998) 3586-3616.
5. Oostenbrink, C., A. Villa, A.E. Mark, W.F. Van Gunsteren, A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6, Journal of Computational Chemistry **25** (2004) 1656-1676.
6. Jorgensen, W.L., J. Tirado-Rives, The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin, Journal of the American Chemical Society **110** (1988) 1657-1666.
7. Kohlhoff, K.J., D. Shukla, M. Lawrenz, G.R. Bowman, D.E. Konerding, D. Belov, R.B. Altman, V.S. Pande, Cloud-Based Simulations on Google Exacycle Reveal Ligand-Modulation of GPCR Activation Pathways, Nature chemistry **6** (2014) 15-21.
8. Korb, O., P.W. Finn, G. Jones, The Cloud and Other New Computational Methods to Improve Molecular Modelling, Expert Opinion on Drug Discovery **9** (2014) 1121-1131.
9. Shaw, D.E., M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Ierardi, I. Kolossváry, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S.C. Wang, Anton, a Special-Purpose

Machine for Molecular Dynamics Simulation, SIGARCH Comput. Archit. News **35** (2007) 1-12.

10. Shaw, D.E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-Level Characterization of the Structural Dynamics of Proteins, Science **330** (2010) 341-346.
11. Lindorff-Larsen, K., S. Piana, R.O. Dror, D.E. Shaw, How Fast-Folding Proteins Fold, Science **334** (2011) 517-520.
12. Blöchliger, N., A. Vitalis, A. Caflisch, A Scalable Algorithm to Order and Annotate Continuous Observations Reveals the Metastable States Visited by Dynamical Systems, Computer Physics Communications **184** (2013) 2446-2453.
13. Blochliger, N., A. Vitalis, A. Caflisch, High-Resolution Visualisation of the States and Pathways Sampled in Molecular Dynamics Simulations, Sci. Rep. **4** (2014)
14. Bleicher, K.H., H.-J. Bohm, K. Muller, A.I. Alanine, Hit and Lead Generation: Beyond High-Throughput Screening, Nat Rev Drug Discov **2** (2003) 369-378.
15. Jorgensen, W.L., The Many Roles of Computation in Drug Discovery, Science **303** (2004) 1813-1818.
16. Shan, Y., E.T. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, D.E. Shaw, How Does a Drug Molecule Find Its Target Binding Site?, Journal of the American Chemical Society **133** (2011) 9181-9183.
17. Seo, M.-H., J. Park, E. Kim, S. Hohng, H.-S. Kim, Protein Conformational Dynamics Dictate the Binding Affinity for a Ligand, Nat Commun **5** (2014)
18. Arora, K., C.L. Brooks, Large-Scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-Shift Mechanism, Proceedings of the National Academy of Sciences **104** (2007) 18496-18501.
19. Boehr, D.D., R. Nussinov, P.E. Wright, The Role of Dynamic Conformational Ensembles in Biomolecular Recognition, Nat Chem Biol **5** (2009) 789-796.
20. Zhou, H.-X., From Induced Fit to Conformational Selection: A Continuum of Binding Mechanism Controlled by the Timescale of Conformational Transitions, Biophysical Journal **98** (2010) L15-L17.
21. Weikl, T.R., F. Paul, Conformational Selection in Protein Binding and Function, Protein Science **23** (2014) 1508-1518.

22. Csermely, P., R. Palotai, R. Nussinov, Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events, Trends in Biochemical Sciences **35** (2010) 539-546.
23. Kolmodin, K., J. Åqvist, Prediction of a Ligand-Induced Conformational Change in the Catalytic Core of Cdc25A, FEBS Letters **465** (2000) 8-11.
24. Boehr, D.D., P.E. Wright, How Do Proteins Interact?, Science **320** (2008) 1429-1430.
25. Buch, I., T. Giorgino, G. De Fabritiis, Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations, Proceedings of the National Academy of Sciences **108** (2011) 10184-10189.
26. Plattner, N., F. Noe, Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models, Nat Commun **6** (2015)
27. Copeland, R.A., D.L. Pompliano, T.D. Meek, Drug-Target Residence Time and Its Implications for Lead Optimization, Nat Rev Drug Discov **5** (2006) 730-739.
28. Swinney, D., Applications of Binding Kinetics to Drug Discovery, Pharmaceutical Medicine **22** (2008) 23-34.
29. Xu, Y., J. Shen, X. Luo, X. Shen, K. Chen, H. Jiang, Steered Molecular Dynamics Simulations of Protein-Ligand Interactions, Science in China Series B: Chemistry **47** (2004) 355-366.
30. Curcio, R., A. Caflisch, E. Paci, Change of the Unbinding Mechanism upon a Mutation: A Molecular Dynamics Study of an Antibody–Hapten Complex, Protein Science: A Publication of the Protein Society **14** (2005) 2499-2514.
31. Patel, J.S., A. Berteotti, S. Ronsisvalle, W. Rocchia, A. Cavalli, Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5, Journal of Chemical Information and Modeling **54** (2014) 470-480.
32. Huang, D., A. Caflisch, Small Molecule Binding to Proteins: Affinity and Binding/Unbinding Dynamics from Atomistic Simulations, ChemMedChem **6** (2011) 1578-1580.
33. Dror, R.O., A.C. Pan, D.H. Arlow, D.W. Borhani, P. Maragakis, Y. Shan, H. Xu, D.E. Shaw, Pathway and Mechanism of Drug Binding to G-protein-Coupled Receptors, Proceedings of the National Academy of Sciences **108** (2011) 13118-13123.
34. Shukla, D., M. Lawrenz, V.S. Pande, *Chapter Twenty-Five - Elucidating Ligand-Modulated Conformational Landscape of GPCRs Using Cloud-*

Computing Approaches, in *Methods in Enzymology*, K.S. Arun, Editor. 2015, Academic Press. p. 551-572.

35. Reymond, J.-L., The Chemical Space Project, Accounts of Chemical Research **48** (2015) 722-730.
36. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and Testing of a General Amber Force Field, Journal of Computational Chemistry **25** (2004) 1157-1174.
37. Vanommeslaeghe, K., A.D. MacKerell, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, Journal of Chemical Information and Modeling **52** (2012) 3144-3154.
38. Vanommeslaeghe, K., E.P. Raman, A.D. MacKerell, Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges, Journal of Chemical Information and Modeling **52** (2012) 3155-3168.
39. Pazos, E., O. Vazquez, J.L. Mascarenas, M. Eugenio Vazquez, Peptide-Based Fluorescent Biosensors, Chemical Society Reviews **38** (2009) 3348-3359.
40. Hao, J., A.W.R. Serohijos, G. Newton, G. Tassone, Z. Wang, D.C. Sgroi, N.V. Dokholyan, J.P. Basilion, Identification and Rational Redesign of Peptide Ligands to CRIP1, A Novel Biomarker for Cancers, PLoS Computational Biology **4** (2008) e1000138.
41. Lee, B.-S., J.-S. Huang, G.D.L. Jayathilaka, S. Lateef, S. Gupta, *Production of Antipeptide Antibodies*, in *Immunoelectron Microscopy*, S.D. Schwartzbach and T. Osafune, Editors. 2010, Humana Press. p. 93-108.
42. Vlieghe, P., V. Lisowski, J. Martinez, M. Khrestchatisky, Synthetic Therapeutic Peptides: Science and Market, Drug Discovery Today **15** (2010) 40-56.
43. Sheinerman, F.B., R. Norel, B. Honig, Electrostatic Aspects of Protein–Protein Interactions, Current Opinion in Structural Biology **10** (2000) 153-159.
44. Suh, J.-Y., C. Tang, G.M. Clore, Role of Electrostatic Interactions in Transient Encounter Complexes in Protein−Protein Association Investigated by Paramagnetic Relaxation Enhancement, Journal of the American Chemical Society **129** (2007) 12954-12955.
45. Harris, B.Z., B.J. Hillier, W.A. Lim, Energetic Determinants of Internal Motif Recognition by PDZ Domains, Biochemistry **40** (2001) 5921-5930.
46. Hillier, B.J., K.S. Christopherson, K.E. Prehoda, D.S. Bredt, W.A. Lim, Unexpected Modes of PDZ Domain Scaffolding Revealed by Structure of nNOS-Syntrophin Complex, Science **284** (1999) 812-815.

47. Christopherson, K.S., B.J. Hillier, W.A. Lim, D.S. Bredt, PSD-95 Assembles a Ternary Complex with the N-methyl-D-aspartic Acid Receptor and a Bivalent Neuronal NO Synthase PDZ Domain, Journal of Biological Chemistry **274** (1999) 27467-27473.
48. Kim, E., M. Sheng, PDZ Domain Proteins of Synapses, Nat Rev Neurosci **5** (2004) 771-781.
49. Saras, J., C.-H. Heldin, PDZ Domains Bind Carboxy-Terminal Sequences of Target Proteins, Trends in Biochemical Sciences **21** (1996) 455-458.
50. Harris, B.Z., F.W. Lau, N. Fujii, R.K. Guy, W.A. Lim, Role of Electrostatic Interactions in PDZ Domain Ligand Recognition, Biochemistry **42** (2003) 2797-2805.
51. Sugita, Y., Y. Okamoto, Replica-Exchange Molecular Dynamics Method for Protein Folding, Chemical Physics Letters **314** (1999) 141-151.
52. Bacci, M., A. Vitalis, A. Caflisch, A Molecular Simulation Protocol to Avoid Sampling Redundancy and Discover New States, Biochimica et Biophysica Acta (BBA) - General Subjects **1850** (2015) 889-902.

Chapter 2

Protein Structural Memory Influences Ligand Binding Mode(s) and Unbinding Rates

Xu, M., Caflisch, A. and Hamm, P., *Journal of Chemical Theory and Computation 2016.*

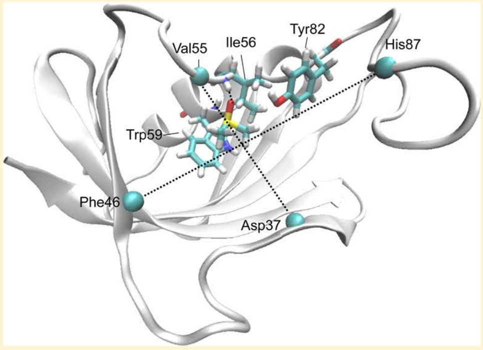
JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

Protein Structural Memory Influences Ligand Binding Mode(s) and Unbinding Rates

Min Xu,[†] Amedeo Caflisch,*[,†] and Peter Hamm*[,‡]

[†]Department of Biochemistry and [‡]Department of Chemistry, University of Zürich, Winterthurerstrasse 190, Zürich CH-8057, Switzerland

Supporting Information



ABSTRACT: The binding of small molecules (e.g., natural ligands, metabolites, and drugs) to proteins governs most biochemical pathways and physiological processes. Here, we use molecular dynamics to investigate the unbinding of dimethyl sulfoxide (DMSO) from two distinct states of a small rotamase enzyme, the FK506-binding protein (FKBP). These states correspond to the FKBP protein relaxed with and without DMSO in the active site. Since the time scale of ligand unbinding (2−20 ns) is faster than protein relaxation (100 ns), a novel methodology is introduced to relax the protein without having to introduce an artificial constraint. The simulation results show that the unbinding time is an order of magnitude longer for dissociation from the DMSO-bound state (holo-relaxed). That is, the actual rate of unbinding depends on the state of the protein, with the protein having a long-lived memory. The rate thus depends on the concentration of the ligand as the apo and holo states reflect low and high concentrations of DMSO, respectively. Moreover, there are multiple binding modes in the apo-relaxed state, while a single binding mode dominates the holo-relaxed state in which DMSO acts as hydrogen bond acceptor from the backbone NH of Ile56, as in the crystal structure of the DMSO/FKBP complex. The solvent relaxes very fast (~1 ns) close to the NH of Ile56 and with the same time scale of the protein far away from the active site. These results have implications for high-throughput docking, which makes use of a rigid structure of the protein target.

1. INTRODUCTION

The association of endogenous ligands to enzymes and receptors regulates a large variety of biochemical pathways. Small-molecule drugs act by specific binding to a target protein to modulate their function, which is usually deregulated in the pathological state. Thus, the detailed analysis of ligand binding and the response of the target protein are expected to give insights useful for the understanding of biochemical processes (e.g., enzymatic reactions and signaling cascades) in the healthy and disease states.[1−4]

Here, we analyze the kinetics and thermodynamics of DMSO unbinding from FKBP by explicit solvent molecular dynamics (MD) simulations. We have selected the 107-residue protein FKBP since there exists a high-resolution crystal structure of the complex with DMSO (PDB code 1D7H; Figure 1).[5] FKBP is a peptidylprolyl isomerase (PPIase), whose name originates from the fact that it binds the immunosuppressant drug FK506 (also called tacrolimus). PPIases are present in many diverse organisms. There are two main classes of PPIases: FKBPs[6−8] and cyclophilins. These two classes are defined in terms of their artificial ligands, the natural products FK506 and cyclosporin A. These macrocyclic immunosuppressants inhibit the rotamase activity of their respective class but have no influence on the other class. Experimental evidence indicates that inhibition of the PPIase activity is not related to immunosuppression.[9] On the other hand, PPIases are important for accelerating the folding of some proteins[10,11] for which the rate-limiting step involves the *trans* to *cis* isomerization of proline peptide bonds.[10−12] Thus, FKBP is involved in protein folding and at the same time is an important small-molecule drug target.

In previous simulation studies, we have shown that unbinding of DMSO from FKBP required about 4 ns at 310 K, and multiple binding modes in the FKBP active site were observed.[13,14] As an extension of that work, we present here multiple independent simulations of spontaneous DMSO unbinding starting from a protein ensemble that has been carefully equilibrated with either a DMSO molecule bound (i.e., in the holo-relaxed state) or in its apo-relaxed state. The simulations show that unbinding from the holo-relaxed state is an order of magnitude slower with the binding mode stabilized by a hydrogen bond with the backbone NH of Ile56. On the other hand, multiple binding modes are observed in the apo-relaxed state. We will furthermore show that upon unbinding from the holo-relaxed state the rearrangement of the FKBP binding site and water network are coupled, that is, both show similar relaxation kinetics. Moreover, the protein and water response is significantly slower than the DMSO unbinding

Received: November 5, 2015

DOI: 10.1021/acs.jctc.5b01052
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

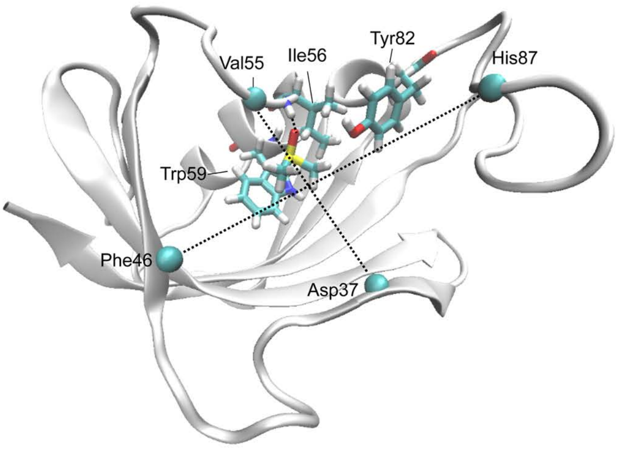


Figure 1. Crystal structure of the complex between FKBP and DMSO (PDB code 1D7H[5]). The DMSO molecule and side chains of Ile56, Trp59, and Tyr82 are shown in sticks. The hydrogen bond of DMSO to the backbone NH of Ile56 is shown by dotted lines. Also shown are the C_α atoms of Phe46, His87, Asp37, and Val55, the distances between which define the deformation of the binding groove after unbinding.

time. The results thus suggest that ligand unbinding is non-Markovian, and the term k_{off} becomes rather meaningless as the rate depends on the state of the protein, which in turn relaxes slower than unbinding itself.

2. METHODS

2.1. MD Setup. All the MD simulations involved in this study were performed using GROMACS 4.6.[15] We used the CHARMM27 force field[16,17] for the protein and CGENFF force field[18] for the small molecule DMSO. The starting conformation was taken from the first chain of the X-ray crystal structure 1D7H (which has a DMSO molecule bound), and the water molecules close to the protein surface were kept. The system was solvated in a rhombic dodecahedron box with ≈7600 additional TIP3P water molecules[19] together with 150 mM NaCl to neutralize the simulation box. We used 2 fs as the integration time step and the LINCS algorithm[20] to constrain the covalent bonds involving hydrogen atoms. Electrostatic interactions were approximated using the Particle-Mesh Ewald (PME) summation method,[21] and van der Waals interactions were truncated with a cutoff of 10 Å. Before relaxation of the protein, the solvent was equilibrated for 1 ns with all the heavy atoms of the complex fixed around their position with a force constant of 1000 kJ/mol nm^2. All simulations were run as an NPT ensemble at a pressure of 1 atm using a Berendsen barostat with coupling time 2 ps and at a temperature of 273 K using velocity rescaling with a stochastic term[22] with coupling time 0.1 ps. The low temperature is necessary to fully equilibrate the DMSO-bound state (see next subsection) because of previous unbinding simulations at 310 K, from which a k_{off}^{-1} of about 4 ns was determined.[13,14] Note that the freezing point of TIP3P water is significantly lower than 273 K in the CHARMM27 force field.[23]

2.2. Relaxation of Bound State. Starting from the solvent-equilibrated state, the goal was to relax the protein in its holo form (i.e., DMSO−protein complex) under the particular force field used in the simulation (which might result in a structure that deviates from the X-ray structure). That was hampered by the fact that the average unbinding time of the unrelaxed protein was k_{off}^{-1} = 15 ns (Figure S1, Supporting Information), which turned out to be significantly too short to relax the protein. We therefore developed the following protocol: The simulation was run until the DMSO−protein distance exceeded some threshold. To that end, we used two distances, the distance r_{Trp59} of the S atom of the DMSO molecule to the center of the second ring of Trp59 as well as the hydrogen bond distance r_{Ile56} of the O atom of DMSO to the amide-proton of Ile56 (Figure 1), and we assumed unbinding when both distances exceeded 12 Å. Once unbinding occurred, we traced back the trajectory until both $r_{\text{Trp59}} < 3.5$ Å and $r_{\text{Ile56}} < 1.9$ Å. At that point of the trajectory, we reassigned new random velocities to all atoms of the simulation box according to a Boltzmann distribution at the same temperature T = 273 K (note that since kinetic and potential energy contributions to the partition function are separate reassigning new velocities and keeping the positions unchanged results in a canonical ensemble). With that initial condition, the MD simulation was launched again until the next unbinding event occurred, which is different and tentatively later than the previous one (Figure 2) since the new run produces a different trajectory. The procedure was continuously repeated.

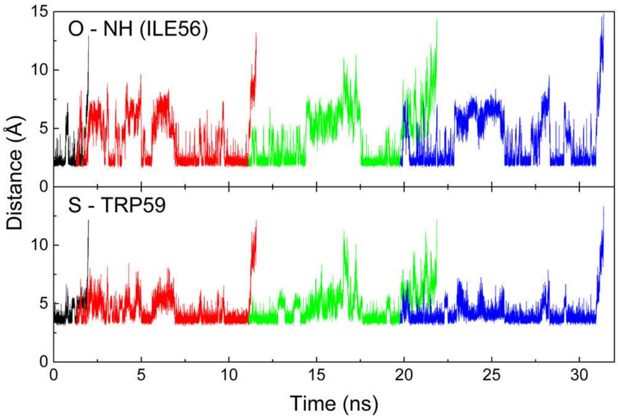


Figure 2. Example of four consecutive trajectories during the relaxation of the bound state, colored in black, red, green, and blue, showing the distances of the O atom of DMSO to the amide-proton of Ile56, r_{Ile56} (top), and that of the S atom of the DMSO molecule to the center of the 6-ring of Trp59, r_{Trp59} (bottom).

Following that protocol, we enable the protein to fully relax with DMSO bound to the protein all the time along that chain of simulations without having to invoke an artificial constraint to force the DMSO molecule to stay in the binding pocket. At the same time, the starting points of the individual simulation pieces serve as an ensemble of bound configurations used as starting points for the nonequilibrium simulations described in the next paragraph. In total, 116 such bound configurations have been collected from two independent chains of trajectories, each ≈1.4 μs long. Bound configurations from the first 200 ns were discarded to ensure that there is no bias from the initial X-ray structure.

2.3. Unbinding Simulations. Starting from the aforementioned 116 bound configurations, trajectories of 150 ns length have been simulated with 1 ps saving time and subsequently 500 ns with 10 ps saving time. By construct, this constitutes an ensemble of nonequilibrium trajectories since all these 116 trajectories start from a bound state, while the equilibrium lies strongly on the unbound side. The k_{on} time for rebinding of the DMSO molecule is in the order of a few microseconds. Hence, out of the 116 trajectories, five such rebinding events occurred after the first 150 ns, and those trajectories have been discarded. For the remaining 111 trajectories, the DMSO has been removed and replaced by three water molecules after 150 ns to prevent rebinding within the subsequent 500 ns. The three water molecules replacing the DMSO molecule were placed at the positions of the oxygen atom and carbon atoms of the methyl groups. The three water molecules were then minimized and equilibrated by a MD run of 1 ns with the rest of the system frozen.

The final trajectories were synchronized by searching for the time point when both distances r_{Trp59} and r_{Ile56} first exceeded 20 Å (a looser criterion was used here to minimize events of immediate rebinding). From that time point, we traced back until both distances are still below 6 Å, which we defined as the time point of unbinding.

2.4. Rebinding Simulations.

In order to also determine the unbinding constant for the apo-relaxed protein, the endpoint configurations of the 111 unbinding simulations described in the previous paragraph have been considered, all of which haven been in the apo state for at least 500 ns, which is long enough to assume that the protein is indeed relaxed. A DMSO molecule has been reintroduced into these endpoint configurations replacing three water molecules, which form a triangle with their length all below 3.6 Å. Furthermore, in order to speed up the time for rebinding, only waters in the vicinity of the binding pocket were considered (i.e., with 13-21 Å from the C_α atoms of Asp37, Phe46, Val55, and His87). These three water molecules have been replaced by the oxygen and carbon atoms of the DMSO molecule with the RMSD minimized. Subsequently, the DMSO molecule as well as all water molecules and ions within 6 Å of the DMSO have been minimized and equilibrated for 1 ns, keeping all other atoms fixed. From that point, 30 ns long simulations were launched with random velocities assigned to all atoms. Typically ≈7% of those simulations rebind during 30 ns, using a threshold of 6 Å for both r_{Trp59} and r_{Ile56} to determine binding. Trajectories that did not bind within 30 ns were relaunched with a new set of random velocities. That procedure was repeated seven times, collecting 58 rebinding events, whose subsequent unbinding was simulated as well. A table listing all simulations used for the study is given in Table S1 of the Supporting Information.

2.5. Solvation Layer.

In order to determine the water layer around the protein after unbinding from the holo-relaxed form, an average structure was first calculated from all snapshots before the time point of unbinding by aligning them upon each other, minimizing the RMSD of all C_α atoms. That average structure was considered to be the reference structure. Snapshot structures after the unbinding event were aligned to that reference structure and averaged in time bins on a logarithmic time axis (i.e., 10 bins per decade). Whenever aligning the protein backbone, the surrounding water molecules and the DMSO molecule were moved accordingly, and their densities have been calculated by binning them into cubes of 1 Å^3. Figure S2 of the Supporting Information shows the RMSD of the structures within one time bin relative to the average in the same time bin, while the point at zero combines all structures before the unbinding event. The RMSD is quite small throughout (0.4 Å, which is consistent with the small B-factors in the 1D7H crystal structure,[5] as 87% of the Cα atoms have B-factor smaller than 30 Å^2) and hardly changes during unbinding and during the relaxation of the protein; hence, the protein remains equally rigid at all times. This in turn evidences that the alignment procedure does not cause artifacts in the calculation of the water density around the protein.

3. RESULTS

3.1. Unbinding Constant and Binding Modes.

Unbinding from the holo-relaxed state is about an order of magnitude slower (k_{off}^{-1} = 23 ns) than from the apo-relaxed state (k_{off}^{-1} = 2.2 ns) (Figure 3). Assuming that binding is diffusion-controlled and hence the same in both cases, the slow down by a factor 10 corresponds to a stabilization of the binding free energy in the holo-relaxed state by ≈ 2.3 k_BT ≈ 1.3 kcal/mol. This difference is significant given that the free energy of binding is only −0.8 kcal/mol as measured by nuclear magnetic resonance spectroscopy[24] or −2.3 kcal/mol measured by tryptophan fluorescence.[5]

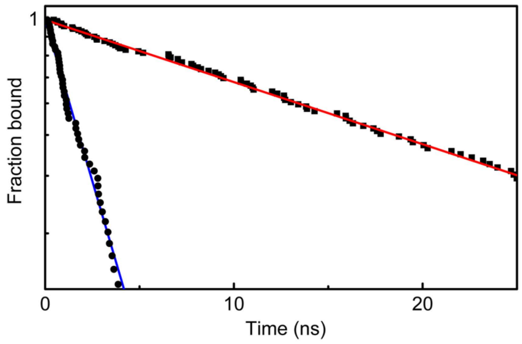


Figure 3. Fraction of bound DMSO molecules as a function of time, using a threshold of 12 Å for both r_{Trp59} and r_{Ile56} as a criterion of unbinding. The solid lines represent exponential fits with unbinding constants k_{off}^{-1} = 23 ns (red) for the holo-relaxed protein and k_{off}^{-1} = 2.2 ns (blue) for the apo-relaxed protein.

Figure 4 shows projections of the free energy onto the two distances r_{Trp59} and r_{Ile56} evaluated along the trajectory segments during which the DMSO molecule is bound. Note that these free energy projections are not rigorously defined since the system is not in equilibrium during these time periods. These two-order parameters can resolve the various binding modes of the DMSO molecule reasonably well. That is, for the

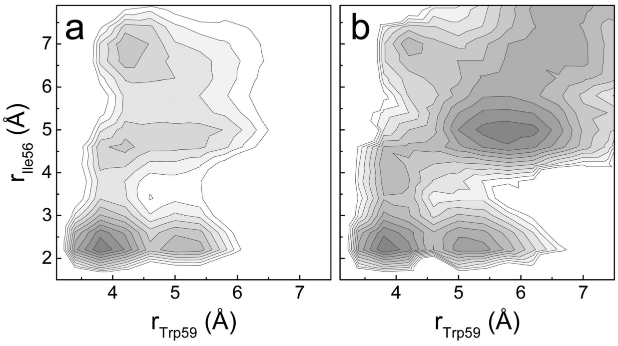


Figure 4. Histogram-based two-dimensional projection of the free energy along the two distances r_{Trp59} and r_{Ile56} for DMSO bound to the holo-relaxed protein (a) and apo-relaxed protein (b). Contour lines are separated by 0.5 k_BT, starting from the corresponding minimum.

holo-relaxed protein (Figure 4a), the most populated binding mode (around $r_{\mathrm{Trp59}} \approx 3.5$ Å and $r_{\mathrm{Ile56}} \approx 2.2$ Å) corresponds to the crystal structure (PDB code 1D7H) in which the oxygen atom of DMSO is involved as acceptor in a hydrogen bond with the NH group of Ile56 (Figure 1). The free energy surface is quite different in the apo-relaxed form (Figure 4b). The minimum with hydrogen bond to the NH of Ile56 still exists but is less populated as only 3 out of 58 trajectories visited that minimum (Methods section). Importantly, a new metastable state is populated ($r_{\mathrm{Trp59}} \approx 5.8$ Å and $r_{\mathrm{Ile56}} \approx 5.0$ Å) in which the oxygen atom of DMSO accepts a hydrogen bond from the hydroxyl group of Tyr82 (see Figure 1 as well as Figure S3, Supporting Information). This state is essentially absent for the holo-relaxed protein. The free-energy difference from the minimum with hydrogen bond to Tyr82 to a fully dissociated state is only $\approx 1.5\ k_{\mathrm{B}}T$, along the "channel" guiding toward the top-right corner of Figure 4b, which explains at least in part the faster unbinding from the apo-relaxed state.

3.2. Slow Response of the Protein. The actual event of unbinding takes roughly a nanosecond (Figure 2). Subsequently, the binding groove relaxes by becoming a bit larger in the one direction (Phe46−His87), and a bit smaller in a perpendicular direction (Asp37−Val55) (Figure 5a). The structural changes are small, that is, $\lesssim 1$ Å, but clearly resolvable. The relaxation process takes a few 100 ns to finish and proceeds in a nonexponential manner. Globally fitting the two distance traces with a stretched-exponential function $S(t) = A\exp(-(t/\tau)^{\beta}) + A_0$ reveals a time constant of $\tau = 65$ ns and a stretching factor of $\beta = 0.77$ that deviates significantly from 1. With these values, we obtain for the mean relaxation time $\langle\tau\rangle \equiv \int_0^{\infty} t\cdot e^{-(t/\tau)\beta}\,\mathrm{d}t = \tau/\beta\cdot\Gamma(1/\beta) = 75$ ns.[25] Importantly, the mean relaxation time is significantly slower than the unbinding constant $k_{\mathrm{off}}^{-1} = 23$ ns.

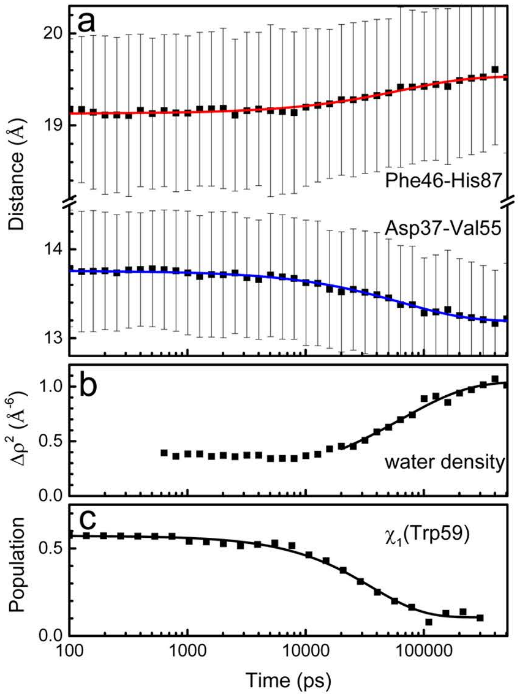


Figure 5. Protein and water relaxation upon DMSO unbinding from the holo-relaxed state. (a) Time dependence of the average $C_{\alpha}-C_{\alpha}$ distance between two pairs of amino acids (Phe46−His87 and Asp37−Val55) across the binding groove. The bars indicate the $\pm\,\sigma$-interval of the fluctuations around the average. (b) Squared water density change $\Delta\rho^2$ integrated around the protein. In panels (a) and (b), the solid lines show a global stretched-exponential fit to the data, revealing a time constant of $\tau = 65$ ns and a stretching factor $\beta = 0.77$. The data <20 ns in panel (b) were discarded, as the initial small drop in $\Delta\rho^2$ is an artifact from the decreasing background noise due to the better averaging in the exponentially increasing time bins. (c) Time dependence of the population of the *gauche*⁻ conformer of the Trp59 χ_1 dihedral angle. The solid line is a single-exponential fit with time constant of 36 ns.

The side chain of Trp59, which forms the "floor" of the DMSO (and substrate) binding site (Figure 1), changes its orientation upon DMSO unbinding with a time scale that is slower than the dissociation but slightly faster than the relaxation of the binding site. The reorientation of the Trp59 side chain, namely, the changes of χ_1 from *gauche*$^-$ to *trans* and χ_2 from about 40° to 0° (Figure S4), is completed in about 100 ns and can be fitted by a single-exponential function with a time constant of 36 ns (Figure 5c). The single-exponential fitting is consistent with a single free energy barrier between the two rotameric states of the Trp59 side chains.

3.3. Solvent Relaxation Coupled to the Protein Response. The temporal evolution of the radial distribution function of the water molecules around the NH of Ile56 (Figure S5a) shows that the rearrangement of the first solvation layer around the Ile56 amide, that is, binding of two to three water molecules, occurs almost simultaneously with the unbinding of DMSO, and is nearly 2 orders of magnitude faster than the full relaxation of the protein. In striking contrast, the relaxation of the solvation layer far away from the DMSO binding site is slow and keeps evolving on a few 100 ns time scale (Figure S6). At about 50 ns, the water density has started to change almost everywhere around the protein, as exemplified for Gly62 and Tyr26 both of which are situated on the protein surface opposite from the binding pocket (Figure S5bc). The change of water density integrated over the whole protein surface shows that this change proceeds essentially simultaneously as the protein undergoes its conformational transition. As a matter of fact, it can be fitted with the same stretched-exponential function with time constant of $\tau = 65$ ns and stretching factor of $\beta = 0.77$ (Figure 5b).

To further investigate the relationship between protein and solvent responses, an additional set of unbinding simulations was carried out with restraints on the protein backbone. In the active site, water still exchanges with DMSO, leading to qualitatively similar results for the water density as in the unrestrained simulations, but essentially nothing happens later on around the protein (Figure S7). These control simulations provide additional evidence that the change in the protein solvation shell and the protein conformation itself are inherently coupled to each other.

4. DISCUSSION AND CONCLUSION

We have carried out multiple MD simulations of DMSO unbinding from two distinct conformations of FKBP, namely, the apo state and DMSO-bound or holo-relaxed state. The MD trajectories were used to extract the unbinding kinetics and analyze the response of the protein and change of the solvent density around it. Four main observations emerge from our simulation study:

(1) Despite the fact that the structural changes of the binding pocket are relatively small ($\lesssim 1$ Å), they have a significant

effect on the DMSO unbinding constant (Figure 3). The differences in the binding site influence the relative weights of the binding modes of DMSO, which are different in the holo-relaxed and apo-relaxed states (Figure 4). The binding mode in the holo-relaxed state is dominated by the direct hydrogen bond to the NH group of Ile56, which is consistent with the X-ray structure (PDB code 1D7H) and previous atomistic simulations carried out at a DMSO concentration of 0.44 M.[14] In the apo-relaxed state, there is a highly populated binding mode in which the DMSO molecule acts as hydrogen bond acceptor for the side chain hydroxyl of Tyr82. This binding mode is kinetically less stable than the one with the hydrogen bond to the NH group of Ile56 (Figure 4).

It is interesting to analyze the mechanism of DMSO (L) binding to FKBP (P), which on the basis of the simulation results can be described by

$$PL \rightleftharpoons P + L \rightleftharpoons P^{b} + L \rightleftharpoons P^{b}L \qquad (1)$$

where P and P^{b} represent the apo and ligand-bound state, respectively. Because the unbinding time from the apo state (2.2 ns) and from the holo state (23 ns) are faster than the relaxation of the protein (middle double arrow, 75 ns), the saturation (Y) of the protein has a hyperbolic dependence on the ligand concentration $[L]$, that is, $Y = [L]/(K_D + [L])$, both at low and high concentrations of the ligand but with different values of the dissociation constant (K_{D}). In other words, the equilibrium is shifted almost completely to the left or right of eq 1 at low or high concentration of DMSO, respectively. Crucially, at intermediate concentrations of the ligand, for example, during a rapid switch in concentration, memory effects will influence the ligand saturation. On the other hand, FKBP is not an hysteretic enzyme in its original definition[26] because the time scale of relaxation upon ligand (DMSO) unbinding (75 ns) is much faster than the reciprocal of the turnover number of the peptidylprolyl isomerase reaction, which is about 10 ms.[27]

(2) The binding/unbinding of DMSO, a small molecule of only four non-hydrogen atoms, has a sizable effect on the protein structure. This observation has important consequences for automatic docking, which is usually performed with a single rigid structure of the protein receptor. In the literature, there are a few examples of high-throughput docking campaigns that made use of multiple protein structures originating from MD simulations and differing from the crystal structure (see ref 28 for a recent review). Crucially, the small-molecule inhibitors identified in those docking campaigns do not fit sterically into the binding site of the original crystal structure.[29−31] In other words, the active compounds discovered in our previous studies, in which the binding site was previously relaxed by MD, would have been false negatives if the docking would have used the rigid crystal structure. Here, we have introduced an MD simulation protocol to "prepare" a protein for docking by relaxation of its complex with a small molecule. The protocol, which consists of the iterative restarting of MD simulations of unbinding, generates a fully relaxed bound state. For the docking of a library of fragments that are similar to the small molecule employed in the equilibration procedure, the use of the equilibrated bound state is more appropriate and will result in less false negatives than the use of the crystal structure of the apo protein. For ligands that have slow unbinding times ($\gtrsim 1\ \mu$s), it is probably sufficient to equilibrate the bound state in a single run starting from the crystal structure of the complex.

(3) With a mean relaxation time of $\langle\tau\rangle = 75$ ns, the rearrangement of the protein and its solvation layer is nearly 2 orders of magnitude slower than the actual process of unbinding ($\approx$1 ns) and also significantly slower than the dissociation time $k_{\mathrm{off}}^{-1} = 23$ ns. This renders the process non-Markovian, and terms like k_{on} and k_{off} become essentially meaningless, as the actual rate depends on the conformational state of the protein (to which it relaxes with the ligand either bound or unbound), which in turn has a long-lived memory. Importantly, the unbinding time becomes a function of the ligand concentration; it is significantly longer at high concentration of DMSO (holo-relaxed) than at low concentration (apo-relaxed).

(4) The first layer of water molecules and atoms on the protein surface relax concomitantly. Thus, the first solvation layer has to be viewed as an integral part of the protein.[32] In other words, the relaxation of the protein surface is not governed by solvent (or vice versa). The relaxation is rather a coupled rearrangement of hydrogen bonds of water molecules and polar groups of the protein (intraprotein and protein−water). The question of the size and properties of the solvation layer has been investigated extensively and discussed in a rather controversial manner over the past years using, for example, NMR spectroscopy,[32] THz spectroscopy,[33] and MD simulations.[34−36] In particular, our results are congruent with the concept of the "slaving" of water and protein motions originally introduced by Frauenfelder, Wolynes, and co-workers.[37,38]

Overall speaking, the response we see here for unbinding of DMSO from FKBP is precisely the same as that for a photoswitchable PDZ2 domain that we have studied recently.[39,40] In that PDZ2 domain, an azobenzene derivative has been covalently linked across its binding groove in a way to mimic a structural change that roughly equals that upon ligand binding/unbinding of the native system. While that molecular construct has been artificial, it has the feature that it can be triggered by light on a picosecond time scale and as such allowed us to study the time response experimentally by transient IR spectroscopy. In ref 39, we obtained a semi-quantitative agreement between the experiments and accompanying MD simulations, which in turn compare well with the MD simulation of the present study for FKBP. That is, we observe the same 100 ns time scale for the response of the binding pocket in both cases, relaxation proceeds in a nonexponential manner, and the size of the structural change is also very comparable with about 1 Å. Also the response of the first solvation layer (Figure S6) is qualitatively the same. This agreement evidences that on the one hand the molecular construct of ref 39 mimics the dynamics of ligand binding reasonably well and on the other hand suggests that the 100 ns time scale is rather general, at least for single-domain proteins of about 100 residues.

It is interesting to relate these results to commonly discussed mechanisms of ligand binding, that is, "induced fit" versus "conformational selection".[41] According to Onsager's regression hypothesis, the relaxation of a nonequilibrium ensemble toward equilibrium equals the equilibrium correlation function, $\langle x(t)\rangle_{\text{nonequ}} \propto \langle x(0)x(t)\rangle_{\text{equ}}$.[42] To that end, it is important to note that the holo-relaxed state effectively constitutes a nonequilibrium ensemble (at low concentration of DMSO), which relaxes toward equilibrium after unbinding of the DMSO molecule. But the size of the structural change on average is of the same order of magnitude as the fluctuations once equilibrium is reached (see bars in Figure 5a); hence, linear response theory applies. We thus can assume that the relaxation kinetics of the opposite process, that is, upon ligand binding, would be the same as that upon ligand unbinding, which we investigated here (just that the DMSO molecule would not stay long enough for full relaxation to occur; Figure 3, blue). In that sense, what we observe here is an "induced fit" mechanism.[41] That is, the protein adapts to a binding event only relatively slowly by lowering the binding free energy and thereby stabilizing the bound state. The question whether it is an induced fit mechanism, however, depends on time scales and concentration of the ligand. DMSO has a very small binding affinity. Hence, its residence time at the protein is short, and the slow relaxation of the protein does indeed matter. On the other hand, for a ligand with much higher binding affinity, the residence time will be significantly longer. Once it exceeds the relaxation time of the protein, the process is effectively Markovian. As the structural change upon ligand binding is smaller than the equilibrium fluctuations, the holo-structure of the protein will occasionally appear even in the apo-relaxed form. The mechanism of ligand binding then turns into "conformational selection".[41]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01052.

Table listing all the simulations performed in this study and figures of the detailed analysis results. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: (+41 44) 635 55 68. Fax: (+41 44) 635 68 62. E-mail: caflisch@bioc.uzh.ch (A.C.).
*Phone: (+41 44) 635 55 68. Fax: (+41 44) 635 68 62. E-mail: peter.hamm@chem.uzh.ch (P.H.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work has been supported by an ERC advanced investigator grant (DYNALLO) and by the Swiss National Science Foundation through the NCCR MUST to P.H., as well as by the Swiss National Science Foundation to A.C. We are grateful for insightful discussions with Andreas Vitalis, Steven Waldauer, and Brigitte Stucki-Buchli.

■ REFERENCES

(1) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Company: New York, 1999.

(2) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964−972.

(3) Seo, M.-H.; Park, J.; Kim, E.; Hohng, S.; Kim, H.-S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **2014**, DOI: 10.1038/ncomms4724.

(4) Plattner, N.; Noe, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.

(5) Burkhard, P.; Taylor, P.; Walkinshaw, M. D. X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies. *J. Mol. Biol.* **2000**, *295*, 953−62.

(6) Siekierka, J. J.; Hung, S. H. Y.; Poe, M.; Lin, C. S.; Sigal, N. H. A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **1989**, *341*, 755−757.

(7) Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. A receptor for the immuno-suppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* **1989**, *341*, 758−760.

(8) Clardy, J. The chemistry of signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 56−61.

(9) Schreiber, S. L.; Crabtree, G. R. The mechanism of action of cyclosporin A and FK506. *Immunol. Today* **1992**, *13*, 136−142.

(10) Fischer, G.; Wittmann-Liebold, B.; Lang, K.; Kiefhaber, T.; Schmid, F. X. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* **1989**, *337*, 476−478.

(11) Tropschug, M.; Wachter, E.; Mayer, S.; Schonbrunner, E. R.; Schmid, F. X. Isolation and sequence of an FK506-binding protein from N. crassa which catalyses protein folding. *Nature* **1990**, *346*, 674−677.

(12) Lang, K.; Schmid, F. X.; Fischer, G. Catalysis of protein folding by prolyl isomerase. *Nature* **1987**, *329*, 268−270.

(13) Huang, D.; Caflisch, A. The free energy landscape of small molecule unbinding. *PLoS Comput. Biol.* **2011**, *7*, e1002002.

(14) Huang, D.; Caflisch, A. Small molecule binding to proteins: Affinity and binding/unbinding dynamics from atomistic simulations. *ChemMedChem* **2011**, *6*, 1578−1580.

(15) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(16) MacKerell, A. D.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(17) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(18) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(20) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(21) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(22) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(23) Vega, C.; Sanz, E.; Abascal, J. L. F. The melting temperature of the most common models of water. *J. Chem. Phys.* **2005**, *122*, 114507.

(24) Dalvit, C.; Floersheim, P.; Zurini, M.; Widmer, A. Use of organic solvents and small molecules for locating binding sites on proteins in solutions. *J. Biomol. NMR* **1999**, *14*, 23−32.

(25) Alvarez, F.; Alegra, A.; Colmenero, J. Relationship between the time-domain Kohlrausch-Williams-Watts and frequency-domain Havriliak-Negami relaxation functions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 7306−7312.

(26) Frieden, C. Kinetic aspects of regulation of metabolic processes: The hysteretic enzyme concept. *J. Biol. Chem.* **1970**, *245*, 5788−5799.

(27) Petros, A. M.; Luly, J. R.; Liang, H.; Fesik, S. W. Conformation of an FK506 analog in aqueous solution is similar to the FKBP-bound conformation of FK506. *J. Am. Chem. Soc.* **1993**, *115*, 9920−9924.

(28) Zhao, H.; Caflisch, A. Molecular dynamics in drug design. *Eur. J. Med. Chem.* **2015**, *91*, 4−14.

(29) Ekonomiuk, D.; Su, X.-C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Otting, G.; Huang, D.; Caflisch, A. Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *J. Med. Chem.* **2009**, *52*, 4860−4868.

(30) Zhao, H.; Caflisch, A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5721−5726.

(31) Zhao, H.; Caflisch, A. Discovery of dual ZAP70 and Syk kinases inhibitors by docking into a rare C-helix-out conformation of Syk. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 1523−1527.

(32) Halle, B. Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc., B* **2004**, *359*, 1207−1224.

(33) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20749−20752.

(34) Brooks, C. L., III; Karplus, M. Solvent effects on protein motion and protein effects on solvent motion. *J. Mol. Biol.* **1989**, *208*, 159−181.

(35) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.* **2000**, *7*, 34−38.

(36) Heyden, M. Resolving anisotropic distributions of correlated vibrations in protein hydration water. *J. Chem. Phys.* **2014**, *141*, 22D509.

(37) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16047−16051.

(38) Lubchenko, V.; Wolynes, P. G.; Frauenfelder, H. The mosaic energy landscapes of liquids and the control of protein conformational dynamics by glass forming solvents. *J. Phys. Chem. B* **2005**, *109*, 7488−7499.

(39) Buchli, B.; Waldauer, S. A.; Walser, R.; Donten, M. L.; Pfister, R.; Blöchliger, N.; Steiner, S.; Caflisch, A.; Zerbe, O.; Hamm, P. Kinetic response of a photoperturbed allosteric protein. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11725−11730.

(40) Waldauer, S. A.; Stucki-Buchli, B.; Frey, L.; Hamm, P. Effect of viscogens on the kinetic response of a photoperturbed allosteric protein. *J. Chem. Phys.* **2014**, *141*, 22D514.

(41) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: An extended view of binding events. *Trends Biochem. Sci.* **2010**, *35*, 539−546.

(42) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: Oxford, 1987.

DOI: 10.1021/acs.jctc.5b01052
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Supplementary Material: Protein structural memory influences ligand binding mode(s) and unbinding rates

Min Xu[1], Amedeo Caflisch[1*], and Peter Hamm[2*]

Department of Biochemistry[1] and Department of Chemistry[2],

University of Zurich, CH-8057 Zurich, Switzerland

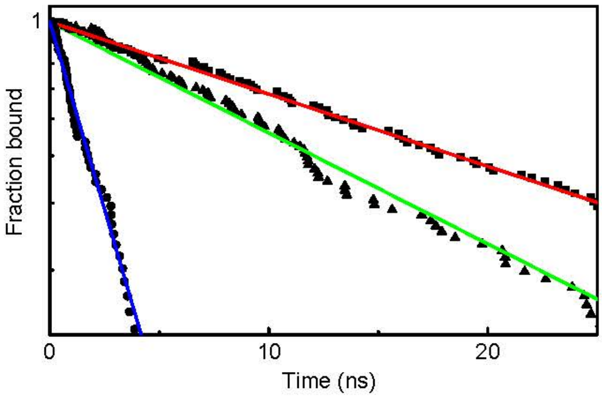


FIG. S1: Fraction of bound DMSO using a threshold of 12 Å. Red and blue lines and corresponding data points are the same as in Fig. 3 of the main text. The triangles and green line originate from unbinding simulations started from the X-ray structure 1D7H (see Methods). The solid lines are exponential fits with unbinding constants $k_{off}^{-1} = 23$ ns for the holo-relaxed protein (red), $k_{off}^{-1} = 15$ ns for the crystal structure (green), and $k_{off}^{-1} = 2.2$ ns for the apo-relaxed protein (blue).

TABLE I: A summary of all simulations performed in this study.

Simulation description	Starting structure	Length of each run	# of independent runs
1. Relaxation, bound state	X-ray	$\approx 1.4\ \mu$s	2
2. Unbinding	ensemble from 1.	150 ns	116
3. Relaxation without DMSO	end structures of 2.	≈ 500 ns	111
4. Rebinding	end structures of 3.	30 ns	58
5. Unbinding, backbone constrained	X-ray	100 ns	100

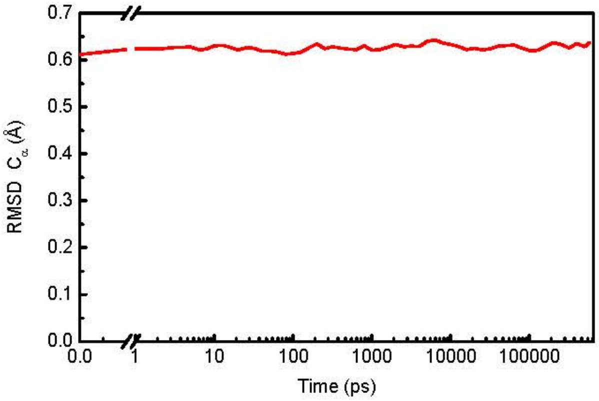


FIG. S2: RMSD of the structures within one time-bin relative to the average structure in the same time-bin. The small RMSD provides evidence that the procedure used for the structural alignment of the MD snapshots for calculating the water density does not introduce artefacts.

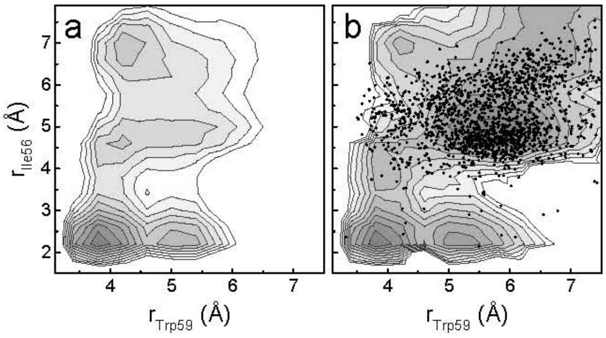


FIG. S3: Same as Fig. 4 in the main text, except for the dots in panel b which correspond to the conformations in which the distance between the DMSO oxygen and the hydroxyl hydrogen atom of Tyr82 is below 2 Å.

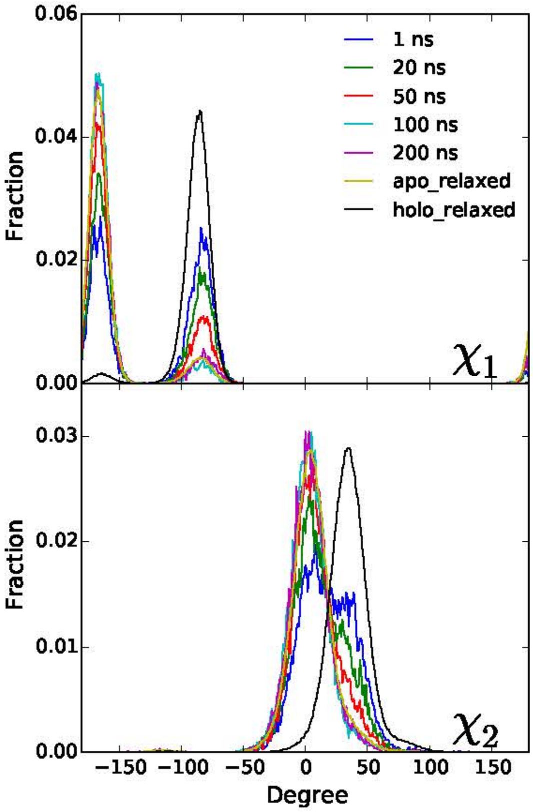


FIG. S4: The distribution of χ_1 and χ_2 dihedral angles of Trp59 shift as a function of time. There are two different conformers of this residue in the apo- and holo-form, respectively. Upon ligand unbinding, there is a progressive shift in the population which converges within 100 ns. The distributions remain nearly constant afterwards.

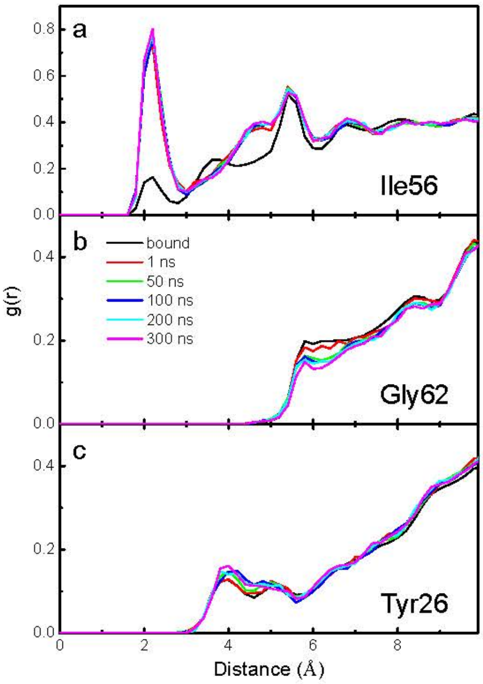


FIG. S5: Radial distribution function $g(r)$ of oxygen atoms of water around the amide proton of Ile56 (a), Gly62 (b), and Tyr26 (c) in the DMSO-bound state (black), as well as 1 ns (red), 50 ns (green), 100 ns (blue), 200 ns (cyan), and 300 ns (magenta) after unbinding from the holo-relaxed protein.

Fig. S5a shows the water structure around the amide proton of Ile56. It is that proton to which DMSO hydrogen-bonds in its most tightly bound configuration (see Fig. 1), and indeed, the corresponding first water peak in the radial distribution function at ≈ 2 Å is small in that case (see Fig. S5a, black line; the peak doesn't vanish completely, since the DMSO molecule also has other non-hydrogen-bonded binding modes). But already 1 ns after unbinding (Fig. S5a, red line), the DMSO molecule is replaced by a hydrogen-bonded water molecule, again as seen from the first peak in the radial distribution function. Water is structured out to ≈ 8 Å, as evidenced by a series of peaks. As a function of time up to 300 ns, the radial distribution function does essentially not evolve any further close to the position of DMSO (Fig. S5a green, blue cyan and magenta lines), while it evolves far away from the binding site (Fig. S5b,c).

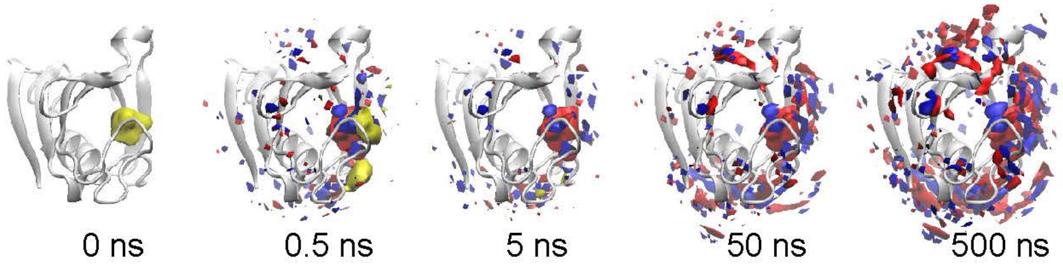


FIG. S6: Change of water density $\Delta\rho$ as a function of time after unbinding from the holo-relaxed protein compared to that just before. Red and blue colors depict increased and decreased water density, respectively. The contour surfaces correspond to changes of ± 0.013 waters/Å^3 (for comparison, the bulk water density is ≈ 0.033 waters/Å^3). The protein is shown by a gray ribbon, and the DMSO density in yellow.

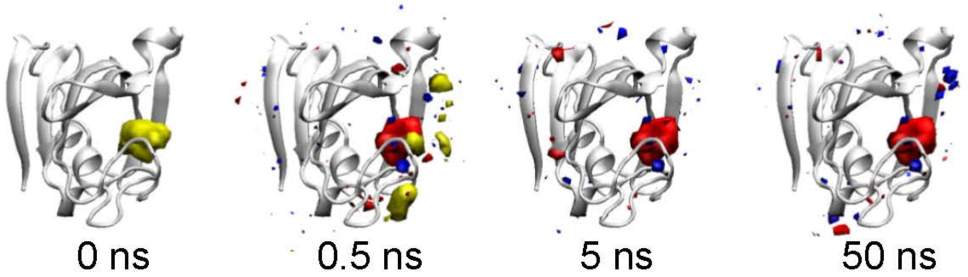


FIG. S7: Same as Fig. S6, but keeping the protein backbone rigid during the unbinding simulations. In this case, a simpler and computationally less expensive protocol has been used to collect an ensemble of starting structures. That is, the DMSO molecule was forced to stay in the binding pocket by center-of-mass pulling with a harmonic spring with a force constant of 1000 kJ/mol·nm^2. The system was then allowed to equilibrate for 100 ns with all other degrees of freedoms free, and 100 snapshots were taken from a subsequent 10 ns run. For the production runs (each of which was 100 ns long), the protein backbone atoms was constrained to the position of the corresponding starting structure with a force constant of 1000 kJ/mol·nm^2. All other procedures were the same as described in Methods.

Chapter 3

Peptide Binding to a PDZ Domain by Electrostatic Steering via Nonnative Salt Bridges

Blöchliger, N., Xu, M. and Caflisch, A., *Biophysical Journal 2015, 108, 9, 2362-2370*

Biophysical Journal Volume 108 May 2015 2362–2370

Article

Peptide Binding to a PDZ Domain by Electrostatic Steering via Nonnative Salt Bridges

Nicolas Blöchliger,[1] Min Xu,[1] and Amedeo Caflisch[1,*]
[1]Department of Biochemistry, University of Zurich, Zurich, Switzerland

ABSTRACT We have captured the binding of a peptide to a PDZ domain by unbiased molecular dynamics simulations. Analysis of the trajectories reveals on-pathway encounter complex formation, which is driven by electrostatic interactions between negatively charged carboxylate groups in the peptide and positively charged side chains surrounding the binding site. In contrast, the final stereospecific complex, which matches the crystal structure, features completely different interactions, namely the burial of the hydrophobic side chain of the peptide C-terminal residue and backbone hydrogen bonds. The simulations show that nonnative salt bridges stabilize kinetically the encounter complex during binding. Unbinding follows the inverse sequence of events with the same nonnative salt bridges in the encounter complex. Thus, in contrast to protein folding, which is driven by native interactions, the binding of charged peptides can be steered by nonnative interactions, which might be a general mechanism, e.g., in the recognition of histone tails by bromodomains.

INTRODUCTION

The fundamental process of protein-protein binding can be conceptualized as diffusional association followed by formation of the stereospecific complex (1–3). Long-range electrostatic forces can significantly accelerate and guide diffusional association, a phenomenon termed electrostatic steering (1,3–9). Association results in a relatively weak encounter complex, which is stabilized mainly by nonspecific interactions and whose binding interface is not yet fully desolvated (1,2,7,10,11). Crossing the transition state, potentially through multiple pathways (12), specific short-range hydrogen bonds and hydrophobic interactions form and the stereospecific complex is reached (1).

PDZ (PSD-95/Discs large/ZO-1) domains, which are found in scaffold proteins involved in signaling (13–15), have been used as model systems to study peptide binding (16,17). They share a common fold with six β-strands and two α-helices (Fig. 1 *A*) and mainly interact with target proteins by binding their C-termini (13,18), although binding to internal protein segments has been reported as well (19). In the stereospecific complex the side chain of a hydrophobic residue at the C-terminus of the target is buried and its carboxylate group interacts with the carboxylate-binding loop. In addition, backbone hydrogen bonds create an intermolecular β-sheet, and specificity is achieved by side-chain interactions (13,20,21).

In a previous molecular dynamics (MD) study on the third PDZ domain of the postsynaptic density protein 95, we focused the analysis on the binding site of the PDZ domain by comparing MD runs of the apo structure of the PDZ domain with MD runs started from the bound state (22). This comparison suggested that the peptide binds by conformational selection. No peptide dissociation event was observed because the length of each of the four trajectories started from the bound state was <0.2 μs. Several other MD simulations of PDZ domains have been performed during the last few years (23–27). However, we are not aware of any MD simulations of the binding of peptides to PDZ domains. Although unbiased MD simulations of sub-μs length have been used already to study the (reversible) binding of small and mainly rigid molecules to proteins (28–32), it is much more challenging to simulate the binding of flexible (oligo)peptides to proteins because the larger conformational space requires significantly longer trajectories (9,12,33).

Here, we report on unbiased, multiple MD simulations of 2.1–3.6 μs each, which were carried out to characterize the binding and unbinding of the C-terminal hexapeptide segment Acetyl-EQVSAV of the Ras-associating guanine nucleotide exchange factor 2 (RA-GEF2, also known as PDZ-GEF2 or RapGEF6) (34) to the second PDZ domain of protein tyrosine phosphatase 1E (PTP1E, also known as PTPL1, FAP-1, or PTP-Bas) (35,36). This study focuses on the intermolecular interactions during the (un)binding process and was motivated by the following questions. Is it possible to capture the spontaneous binding of a flexible hexapeptide to the PDZ domain by MD simulations on a conventional compute cluster? What is the role of the electrostatic interactions in the initial association and final binding? Does the binding proceed through native interactions, i.e., via the intermolecular contacts of the stereospecific complex as observed in the crystal structure? Is unbinding the reverse of binding?

Submitted February 20, 2015, and accepted for publication March 17, 2015.

*Correspondence: caflisch@bioc.uzh.ch

Editor: Rohit Pappu.

0006-3495/15/05/2362/9 $2.00

http://dx.doi.org/10.1016/j.bpj.2015.03.038

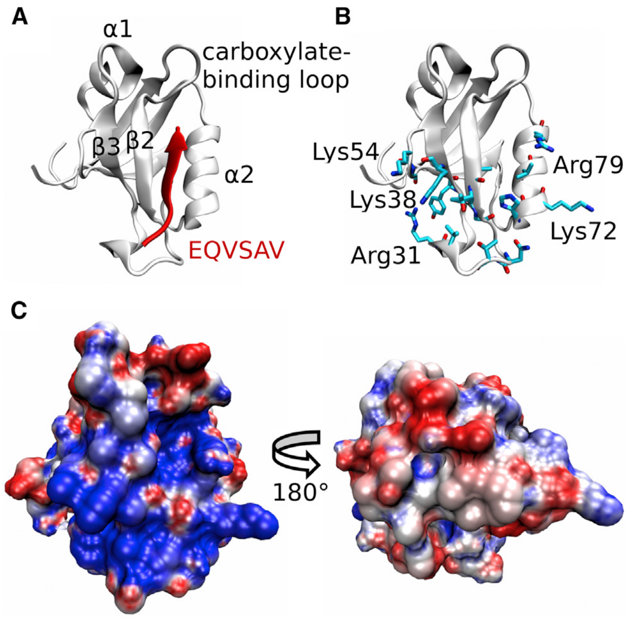


FIGURE 1 Initial association of the peptide and the PDZ domain by electrostatic steering. (*A*) Crystal structure of the stereospecific complex of PTP1E PDZ2 and the C-terminal RA-GEF2 peptide (PDB code 3LNY). The PDZ domain is shown in white with some secondary structure elements labeled. The peptide and its sequence are in red. (*B*) Surface of initial association. PDZ residues having an average contact frequency with the peptide residues of at least 0.1 during association are shown in a stick-like representation (see Materials and Methods). The structure and the orientation are the same as in (*A*). The contact frequency values are shown in Fig. S1. (*C*) Electrostatic surface potential. The color scale ranges from -5 kT/e (*red*) to 5 kT/e (*blue*). The orientation is the same as in (*A*) for the left panel. All illustrations were rendered with VMD (85). To see this figure in color, go online.

We observed multiple events of spontaneous binding in 10 MD runs of ~2.3 μs each started from fully unbound, and several rebinding events in 10 MD runs of ~3.4 μs each started from the crystal structure of the complex. Fast initial association is always driven and stabilized by long-range electrostatic interactions between negatively charged carboxylate groups in the peptide and positively charged side chains in the vicinity of the binding site. These salt bridges are not present in the final stereospecific complex.

MATERIALS AND METHODS

MD simulations

We carried out 10 independent simulations starting with the peptide placed randomly in the simulation box (called binding runs in the following) and 10 simulations started from the bound state (called unbinding runs). The total simulation length amounted to 57 μs.

The coordinates of PTP1E PDZ2 in complex with the C-terminal RA-GEF2 peptide were downloaded from the protein database (Protein Data Bank (PDB) code 3LNY, URL www.rcsb.org) (37). The sequence of the C-terminal RE-GEF2 peptide used here is EQVSAV, and its N-terminus was capped with acetyl. To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and the histidine side chains were neutral. The structure was solvated in a cubic water box. For the binding runs the peptide was placed randomly in the box with a resulting mean distance to the PDZ domain of 12 Å. The size of the box was 73 Å for the binding runs and 63 Å for the unbinding runs. The simulation system contained sodium and chloride ions to approximate an ionic strength of 150 mM and to compensate for the total charge of the two molecules. The simulations were carried out with GROMACS 4.5.5 (38) using the CHARMM27 force field (39,40) and the TIP3P water model (41). Periodic boundary conditions were applied, and electrostatic interactions were evaluated using the particle-mesh Ewald summation method (42). The van der Waals interactions were truncated at a cutoff of 10 Å. The temperature of 310 K was kept constant by an external bath with velocity rescaling (43), and the pressure was kept close to 1 atm by the Berendsen barostat (44). The LINCS algorithm was used to fix the covalent bonds involving hydrogen atoms (45). The integration time step was 2 fs, and snapshots were saved every 10 ps. Each MD run was carried out on 16 cores (i.e., four Xeon5560 CPUs) of the Schrödinger supercomputer at the University of Zurich, which required ~1 week per μs.

SAPPHIRE plot

Recently, we have developed an algorithm for the analysis of long MD trajectories (46,47). The resulting SAPPHIRE (States And Pathways Projected with HIgh REsolution) plot is a comprehensive visualization of the thermodynamics and kinetics of the simulated system. A function measuring distance between snapshots is needed to generate SAPPHIRE plots and can be freely chosen by the user. We chose the Euclidean distance function on 29 distances between atoms of the peptide and the binding site of the PDZ domain for the present application. Table S1 in the Supporting Material contains the full list of atom pairs used.

We briefly describe the method here and refer the reader to the original publications for more details (46,47). Starting from an arbitrary snapshot, all the snapshots are sequentially ordered in a stepwise fashion. In each step, the snapshot closest to any snapshot prior in the sequence becomes the next entry. The complete sequence of snapshots is called progress index. Assuming high snapshot density within free energy basins, snapshots belonging to the same basin are grouped together and distinct states do not overlap (46). A stochastic algorithm to generate an approximate progress index has been developed. This algorithm is scalable to large data sets and was used here. It is important to note that the progress index is not a reaction coordinate. It is rather a sorting of all MD snapshots to identify basins without any a priori clustering.

We employ three types of annotation functions to highlight and interpret the states along the progress index and the pathways connecting them (Fig. 2). First, we use a kinetic annotation function to localize the individual states on the progress index. Specifically, for every snapshot i along the progress index, we plot the average of the mean first-passage times between A_i and S_i, denoted τ_{MFP}, where A_i is the set of snapshots added to the progress index before i and S_i is the set of those added after i. The value of this annotation function is low within a state and high in transition regions, and barriers are highlighted reliably (although they cannot be interpreted quantitatively) (46). Second, we plot the actual sampling time of the individual snapshots to illustrate when and in which sequence the different states were sampled. This information appears as red dots in Fig. 2 and corresponds to the trace of the temporal evolution of the system, i.e., the detailed sequence of events for each MD run. Third, we characterize the states themselves by a structural annotation. In this case we have used the distance between the peptide and the PDZ domain, the solvent accessible surface area of the peptide, the root mean-square deviation (RMSD) of the peptide with respect to a reference structure after alignment on the PDZ domain, as well as several interatomic distances.

Trajectories from the individual simulation runs were concatenated and subsampled at 20 ps to generate the SAPHHIRE plot. For the unbinding runs, the size of the simulation box was adjusted to match the binding runs after the system has been centered on the PDZ atoms. The stochastic

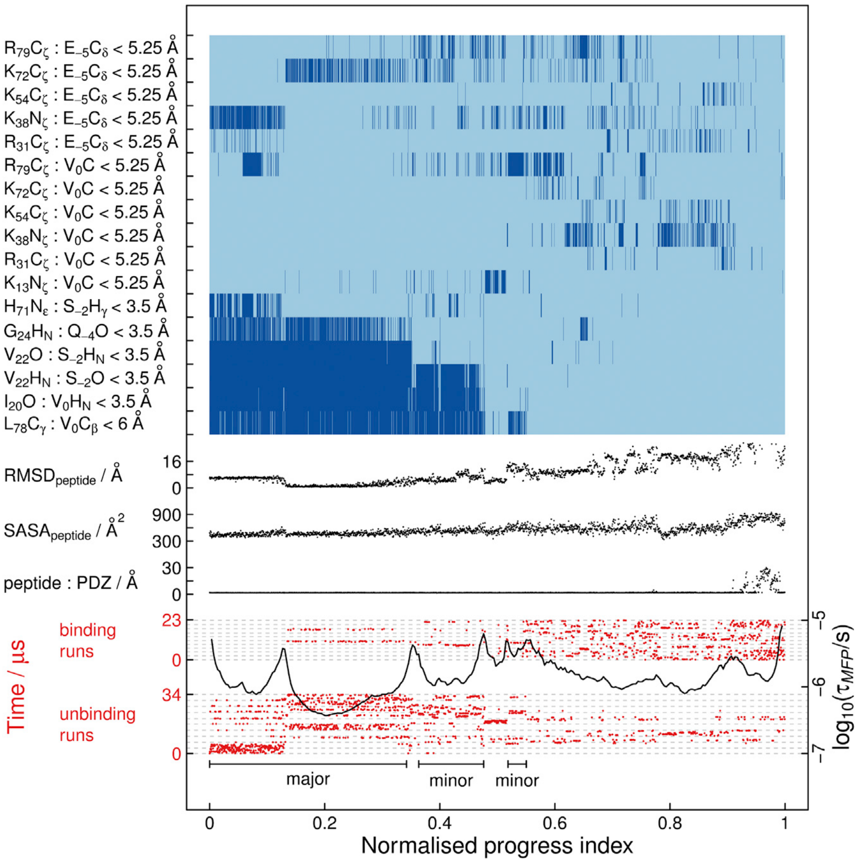


FIGURE 2 SAPPHIRE plot illustrating the sampled ensemble. The progress index (x axis) represents a reordering of the trajectory snapshots that groups similar snapshots next to each other (see Materials and Methods). The progress index is annotated with kinetic information (τ_{MFP}, a function whose value is low within states and high in transition regions, *black profile* in the *bottom*), sampling time (*red dots*), and structural information (*middle* and *top*). The annotation in the top part of the panel uses binning, with dark blue meaning that the distance given on the left side (reporting on burial of the Val0 side chain, intermolecular backbone hydrogen bonds, and salt bridges) is below the indicated threshold. $RMSD_{peptide}$ was computed on the C_α atoms of the peptide with respect to a representative structure of the major binding mode after alignment on the PDZ domain. $SASA_{peptide}$ is the solvent accessible surface area of the peptide. Peptide: PDZ denotes the minimal distance between the peptide and the PDZ domain. Gray dashed lines indicate the boundaries between individual simulation runs. The major and minor binding modes of the stereospecific complex are labeled (*black horizontal segments* in the *bottom*). To see this figure in color, go online.

algorithm mentioned previously is scalable because of the preorganization of the data via tree-based, hierarchical clustering (48). The lower and upper threshold radius and the tree height for the clustering were set to 0.6 Å, 10 Å, and 12. The first snapshot on the progress index is the starting structure of the first run, i.e., the crystal structure of the bound complex (49). The number of guesses to find nearest neighbors (46) was set to 10^4. The method is implemented in the CAMPARI simulation and analysis package (http://campari.sourceforge.net).

Contact frequencies

First, the MD trajectory segments were classified as stereospecific complex or other, where other includes fully unbound and encounter complex. For this binary classification the kinetic annotation of the SAPPHIRE plot was used (*black profile* in Fig. 2), as well as the RMSD of the peptide with respect to the major binding mode, and various distances between the peptide and the binding pocket (Figs. S2–S21). This classification is illustrated in the top of Figs. S2–S21. The 10 binding runs were then used to compute contact frequencies by employing only the segments annotated as other (i.e., fully unbound and encounter complex). A contact is considered to be formed between a residue of the peptide and a residue of the PDZ domain if two atoms are within 5 Å. The acetyl at the N-terminus of the peptide was considered as an independent residue, and CAMPARI (http://campari.sourceforge.net) was used for this analysis.

Electrostatic surface potential

The electrostatic potential on the surface of the PDZ domain was calculated with PDB2PQR (50,51) and APBS (52) using the conformation of the PDZ domain in the stereospecific complex (PDB entry 3LNY).

Binding time and k_{on}

Mean binding times were separately estimated for the binding runs and the rebinding events observed in the unbinding runs as $\tau = t_{unbound}/n$, where $t_{unbound}$ is the total time the peptide is not bound as in the stereospecific complex (defined previously and annotated in Figs. S2–S21) and n is the number of binding events. For the binding runs $n = 5$ and $\tau = 3.8$ μs. We observed $n = 3$ rebinding events in the unbinding runs, resulting in $\tau = 2.4$ μs. The rate constant k_{on} was estimated to be $1/\tau$ [peptide], where [peptide] = 4.3 mM and 6.7 mM for the binding and unbinding runs, respectively. The resulting values for k_{on} are 61 $\mu M^{-1}s^{-1}$ and 63 $\mu M^{-1}s^{-1}$ for the binding and unbinding runs, respectively.

Free energy profile

Cut-based (53,54) and conventional, histogram-based free energy profiles were computed using Fep1d (55).

RESULTS AND DISCUSSION

We performed 10 simulations starting from the peptide placed randomly in the simulation box. These 10 simulations (called binding runs in the following) were completely agnostic of the binding site, and no biasing force or restraint was used. In addition 10 independent runs were started from the bound state using the crystal structure of the complex (PDB code 3LNY (49)) as starting conformation and different random seeds for the initial assignment of the velocities. Total simulation time amounted to 57 μs.

Association via electrostatic steering and nonnative salt bridges

We observed fast association of the peptide and the PDZ domain in all of the binding runs. The intermolecular distance dropped below 2.5 Å within 5 ns on average. The residues most involved in complex formation are Val22, Thr23, Gly24, His71, Val75, and Arg79 located along the binding site, Asn27, Thr28, Val30, Arg31, Tyr36, and Lys38 in the β2-β3 loop and on strand β3, as well as Lys54, Gly55, and Lys72 (Figs. 1 *B* and S1 and Materials and Methods). On the other hand, contact frequencies are low for the carboxylate binding loop, for helix α1, which contains two negatively charged residues, and the β-sheet formed by the β1, β6, β4, and β5 strands, which is located on the other side of the domain with respect to the binding site. The electrostatic potential on this surface of initial association is positively charged (Fig. 1 *C*), and diffusion of the peptide, which bears two negative charges, to the vicinity of the binding site is thus mainly driven by electrostatic steering.

Various salt bridges are formed in the encounter complex, which features multiple relative orientations of the peptide and PDZ domain. In the fifth binding run, for example, the carboxylate group of Val0 (peptide residues are numbered from −5 to 0) forms salt bridges with Arg79, Lys13, and Lys72 before committing to the final binding pose (Fig. 3 and Movie S1). The detailed sequence of events and the roles played by the individual charged residues are different in the other simulation runs (Figs. S2–S21), illustrating the heterogeneity of the encounter complex and the lack of specific interactions (10,11,56–58). However, in all of the binding runs salt bridges are dynamically formed in the encounter complex.

The solvent accessible surface area of the peptide is larger in the encounter complex than in the stereospecific complex (Figs. 3 and S2–S21). This indicates that the binding interface of the encounter complex is not fully desolvated and that specific intermolecular hydrogen bonds or hydrophobic interactions are of lesser importance (10,56,57).

Final binding in antiparallel β-sheet arrangement

We have recently developed a method for the visualization of long MD trajectories (46,47). The main output of the method is the SAPPHIRE plot, which offers an intuitive illustration of the states and sequence of events encountered during the simulation (see Materials and Methods). Previously, we used SAPPHIRE plots to analyze protein folding and conformational changes in the native state of a protein (47) as well as multiple conformations of a loop of the prion protein (59). Here, we apply the method to a binding process.

The SAPPHIRE plot of the combined binding and unbinding runs (Fig. 2) shows that the major binding mode is stabilized by the canonical burial of the Val0 side chain (formed contact with side chain of Leu78 in the binding pocket). In addition, the Val0 carboxylate group interacts with the carboxylate-binding loop (Fig. 3 *A*) and intermolecular backbone hydrogen bonds are formed between the carbonyl

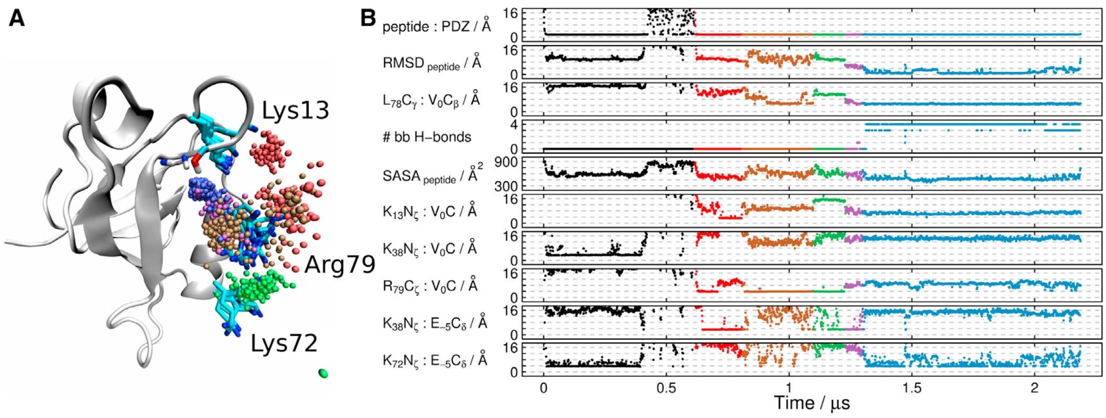


FIGURE 3 Salt bridges stabilize the encounter complex. (*A*) Nonspecific salt bridges in the encounter complex during binding in the fifth binding run (Movie S1). A representative conformation of the PDZ domain in the major binding mode is shown along with the backbone amide groups of the carboxylate-binding loop and the side chain of Ser17. A sphere is drawn every 1 ns at the positon of the carbon atom of the Val0 carboxylate group during binding and colored according to time, as indicated in (*B*). The side chains of selected basic residues involved in salt bridges with the carboxylate group of Val0 are drawn every 150 ns. The illustration was rendered with VMD (85). (*B*) Analysis of the fifth binding run. The minimal distance between the peptide and the PDZ domain (peptide: PDZ), the RMSD of the peptide C_α atoms after alignment on the PDZ domain, the number of backbone hydrogen bonds formed (i.e., distances between the carbonyl oxygen of Ile20 and the NH group of Val0, between both polar groups of Val22 and Ser-2, and between the NH group of Gly24 and the carbonyl oxygen of Gln-4 below 3.5 Å), the solvent accessible surface area of the peptide ($SASA_{peptide}$), and distances between selected atom pairs are plotted as median values in a window of 1 ns. Corresponding plots for the other simulation runs are given in Figs. S2–S21. To see this figure in color, go online.

oxygen of Ile20 and the NH group of Val0, between both polar groups of Val22 and Ser-2, and between the NH group of Gly24 and the carbonyl oxygen of Gln-4 (Fig. 2). Of importance, the most populated binding mode is essentially identical to the crystal structure. The barrier at a value of the normalized progress index of ~0.15 is due to reorientation of the Glu-5 side chain, which can either point toward the solvent or form a salt bridge with Lys72. Regarding the crystal structure, note that the atoms of the Glu-5 residue had very high B-factors and the side chain did not show any electron density (49). Furthermore, the peptide used by Zhang et al. is slightly longer than the one we simulated and has an additional charged residue (Glu-7), which is likely to affect, at least in part, the orientation of the N-terminal segment of the peptide in the bound conformation.

A minor binding mode is located between normalized progress index values of ~0.36 and ~0.48 (Fig. 2). In this binding mode the C-terminal part of the peptide is bound as in the crystal structure, whereas the N-terminal segment protrudes into the solvent. Only the two backbone hydrogen bonds toward the C-terminus of the peptide (between the carbonyl oxygen of Ile20 and the NH group of Val0 and between the NH group of Val22 and the carbonyl oxygen of Ser-2) are formed, in agreement with recent experimental results obtained by amide-to-ester mutations (57). Another alternative binding mode, which is short-lived and was repeatedly sampled, is located between normalized progress index values of ~0.52 and ~0.55. This binding mode features burial of the Val0 side chain (as in the crystal structure), whereas the Val0 carboxylate group forms a salt bridge with Arg79 instead of interacting with the carboxylate-binding loop and no intermolecular backbone hydrogen bonds are present. Finally, snapshots representing the encounter complex are found between normalized progress index values of ~0.55 and ~0.9. Fully unbound conformations accumulate at the end of the progress index.

The stereospecific complex (major or minor binding modes) was reached in five out of the 10 binding runs (Figs. 2, 3, and S2–S11). Additionally, the peptide rebound in three of the six unbinding runs in which full dissociation was observed (Figs. S12–S21 and Movie S2). Our estimate for k_{on} based on these eight binding events is ~60 $\mu\mathrm{M}^{-1}\mathrm{s}^{-1}$ (see Materials and Methods). We note that the TIP3P water model used here shows a self-diffusion constant higher by a factor of 2–3 than the experimentally measured value (60), which might influence k_{on}. Experimental values for k_{on} collected at lower temperatures and similar or higher ionic strength range from 2.9 to 36 $\mu\mathrm{M}^{-1}\mathrm{s}^{-1}$ for the same PDZ domain or its mouse ortholog PTP-BL PDZ2 and the peptide ENEQVSAV or *dansyl*-EQVSAV (49,57,61–63).

The dissociation of the encounter complex is frequent on the timescale of binding in our simulations as the average lifetime of the encounter complex is ~200 ns (see distance between peptide and PDZ domain in Figs. S2–S21). The encounter complex is thus located before the rate-limiting step (2,56,57). This observation is validated by the free energy profile along the distance between the Val0 side chain and the hydrophobic pocket of the PDZ domain (Fig. 4), which confirms that the main barrier accounts for the burial of the Val0 side chain. Comparing Fig. 3 with the corresponding figures for the other simulation runs (Figs. S2–S21) shows that the stereospecific complex can be reached from the encounter complex via various pathways. The burial of the Val0 side chain takes place before the formation of the backbone hydrogen bonds or almost simultaneously (e.g., in the binding run 8, Fig. S19). Thus, the sequence of events for binding starts with the formation of nonnative salt bridges in the encounter complex (which does not always lead to full binding) followed by burial of the Val0 side chain, and formation of the backbone hydrogen bonds between residues Val0/Ser-2 and the PDZ $\beta 2$ strand in an antiparallel β-sheet arrangement.

Inverse sequence of events during unbinding

It is interesting to analyze the unbinding process and compare with binding. Peptide dissociation starts by the rupture of the backbone hydrogen bonds, which takes place before the Val0 side chain exits from the hydrophobic pocket of the PDZ domain. Thus, the initial events of unbinding are the reverse of the final events of binding. Furthermore, the peptide does not immediately diffuse away from the PDZ domain after the native interactions of the stereospecific complex break apart. Instead, the peptide remains in contact with the PDZ domain for several hundred nanoseconds (Movie S2 and Figs. S4, S5, S7–S9, and S21). Quantitatively, the residence time in the encounter complex is 650 ± 900 ns during unbinding and 200 ± 300 ns during binding. Of importance, the same nonnative salt bridges provide kinetic stabilization to the encounter complex during

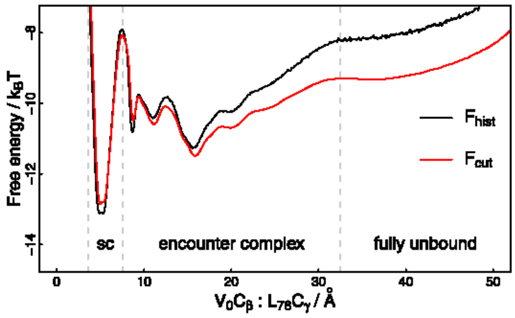


FIGURE 4 Free energy profile along a geometric order parameter. Histogram-based (*black*) and cut-based (*red*) (53,54) free energy profiles are shown as a function of the distance between the C_β atom of Val0 and the C_γ atom of Leu78, which reports on burial of the Val0 side chain. Barriers separating the stereospecific complex (sc), the encounter complex, and fully unbound conformations are indicated by gray, dashed lines. Note that this simple projection introduces overlap and hides crucial information, which, in contrast, is fully resolved by the SAPPHIRE plot (Fig. 2), e.g., the presence of minor binding modes. To see this figure in color, go online.

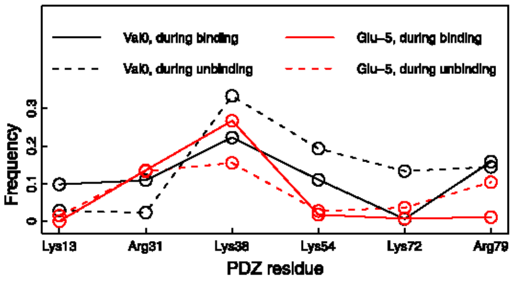


FIGURE 5 Salt bridges in the encounter complex during binding and unbinding. The trajectory segments that correspond to the encounter complex during binding and unbinding were extracted based on Figs. S2–S21, and salt bridges were considered to be formed if the N_ζ atom (for the PDZ lysines) or the C_δ atom (for the PDZ arginines) was within 6 Å of the carboxylate carbon of Val0 or Glu-5, respectively. To see this figure in color, go online.

both peptide association and dissociation (Fig. 5). Thus, the sequence of events for full dissociation is the reverse of binding.

CONCLUSIONS

We have used unbiased MD simulations to analyze the binding of the C-terminal hexapeptide segment of a natural ligand to the second PDZ domain of PTP1E. The general view of the binding process is schematically depicted in Fig. 6 and a representative binding event is shown in Movie S1. Initial association is driven by the long-range electrostatic interactions between the peptide and the PDZ domain (Fig. 1 *C*). In the resulting encounter complex the peptide is weakly bound in the vicinity of the binding site (Figs. 1 *B* and 3). The complex is maintained by nonspecific electrostatic interactions, which allows the peptide to sample multiple orientations (Fig. 3). After the rate-limiting step the side chain of Val0 is buried in a hydrophobic pocket (Figs. 2 and 3). At this point, up to four backbone hydrogen bonds between the peptide and $\beta 2$ can form depending on whether the major binding mode is reached directly or via distinct minor binding modes (Fig. 2). The comparison of the sequence of events for binding and unbinding shows that the two processes are one the inverse of the other.

To further investigate the influence of the encounter complex on the rate constant for binding, we suggest to measure experimentally the salt dependence of the binding rate, e.g., by the Förster resonance energy transfer technique. These measurements have already been reported for PTP-BL PDZ2 and a dansylated peptide without any charged side chains (64). Whereas k_{off} was independent of the ionic strength, k_{on} decreased with increasing ionic strength, which was attributed to the negative charge of the C-terminal carboxylate group. A stronger influence on k_{on} is predicted, on the basis of our MD simulation results, for a similar peptide with one or two negatively charged side chains. On the other hand, electrostatic steering has been ruled out for binding of a peptide with no net charge (*dansyl*-KQTSV) to PDZ3 of postsynaptic density protein 95 (which has glutamic acids at the positions of Arg31 and Lys72) (65).

The mechanism of initial association guided by nonspecific electrostatic steering is likely to be valid for other (small, single-domain) peptide-binding proteins (4,7,9,66,67). As an example, the binding of histone tails to bromodomains is most probably driven by the negative electrostatic potential on the surface surrounding the acetylated lysine binding site, whereas the final stereospecific complex is stabilized by the hydrogen bond between the acetyl carbonyl and the side chain of the evolutionary conserved Asn (68). Other examples include the binding of phosphorylated peptides to SH2 domains (12) as well as intrinsically disordered proteins (69), which tend to contain more charged residues than globular proteins (70). Regarding the coupled binding and folding of intrinsically disordered proteins (71), experimental and theoretical (72,73) studies have highlighted nonnative salt bridges in the encounter complex (74), enhanced on-rates due to electrostatic interactions (75–77), nonnative steering (78), and late formation of native contacts (79).

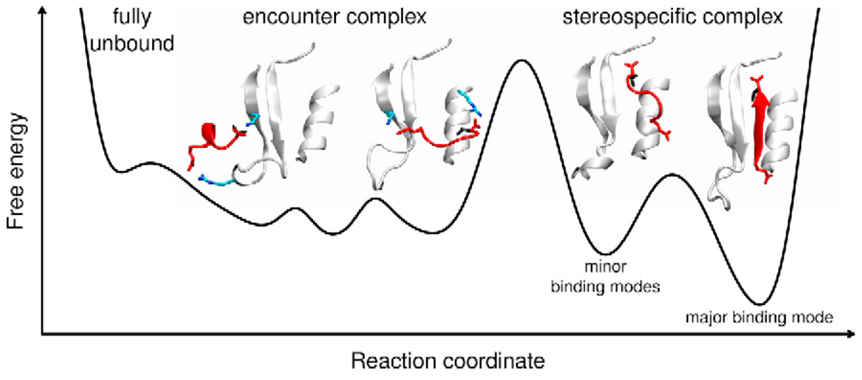


FIGURE 6 Schematic free energy profile of the binding process. After association accelerated by electrostatic steering a weak encounter complex is formed, which is stabilized by nonspecific intermolecular salt bridges. In contrast, the stereospecific complex features burial of the Val0 side chain and multiple binding modes differing among each other only in the orientation of the N-terminal part of the peptide. In this qualitative illustration, relative barrier heights roughly reflect the kinetics observed in the MD simulations, i.e., fast formation, reconfiguration and dissociation of the encounter complex, interconversions among major and minor binding modes on an intermediate timescale, and slow transitions between encounter complex and stereospecific complex. (*Insets*) The ribbon illustrations focus on the binding site, i.e., only the following structural elements are shown for clarity: Carboxylate-binding loop, $\beta 2$-$\beta 3$ strands and loop, and helix $\alpha 2$ of the PDZ domain (*gray*), backbone of the peptide, C-terminal carboxylate group, and the Glu-5 side (*red*), Val0 side chain (*black*), and the side chains of basic residues of the PDZ domain involved in salt bridges with the peptide (*cyan*). The ribbon illustrations were prepared with VMD (85). To see this figure in color, go online.

Finally, it is interesting to compare protein folding with peptide binding as they differ in the number of molecules involved but they are both governed by noncovalent interactions. Protein folding is driven by progressive formation of native interactions, which are in general more favorable than nonnative contacts (80–84). In contrast, our simulation results provide evidence that the binding of a charged peptide to a protein surface with opposite charge can be steered by long-range polar interactions that are not present in the final bound state.

SUPPORTING MATERIAL

Twenty-one figures, one table, and two movies are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00299-4.

AUTHOR CONTRIBUTIONS

All authors contributed to the study design. M.X. performed the MD simulations and prepared the Supplementary Movies, N.B. and A.C. analyzed the data and wrote the article.

ACKNOWLEDGMENTS

This work was supported by a grant from the Swiss National Science Foundation to A.C.

REFERENCES

1. Schreiber, G., G. Haran, and H. X. Zhou. 2009. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* 109:839–860.
2. Schreiber, G. 2002. Kinetic studies of protein-protein interactions. *Curr. Opin. Struct. Biol.* 12:41–47.
3. Berg, O. G., and P. H. von Hippel. 1985. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* 14:131–160.
4. Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3:427–431.
5. Selzer, T., S. Albeck, and G. Schreiber. 2000. Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* 7:537–541.
6. Hemsath, L., R. Dvorsky, …, M. R. Ahmadian. 2005. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Mol. Cell.* 20:313–324.
7. Northrup, S. H., J. O. Boles, and J. C. L. Reynolds. 1988. Brownian dynamics of cytochrome *c* and cytochrome *c* peroxidase association. *Science.* 241:67–70.
8. Gabdoulline, R. R., and R. C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.
9. Ahmad, M., W. Gu, and V. Helms. 2008. Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem. Int.Ed. Engl.* 47:7626–7630.
10. Tang, C., J. Iwahara, and G. M. Clore. 2006. Visualization of transient encounter complexes in protein-protein association. *Nature.* 444:383–386.
11. Suh, J. Y., C. Tang, and G. M. Clore. 2007. Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* 129:12954–12955.
12. Giorgino, T., I. Buch, and G. De Fabritiis. 2012. Visualizing the induced binding of SH2-phosphopeptide. *J. Chem. Theory Comput.* 8:1171–1175.
13. Sheng, M., and C. Sala. 2001. PDZ domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* 24:1–29.
14. Harris, B. Z., and W. A. Lim. 2001. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* 114:3219–3231.
15. Feng, W., and M. Zhang. 2009. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nat. Rev. Neurosci.* 10:87–99.
16. Jemth, P., and S. Gianni. 2007. PDZ domains: folding and binding. *Biochemistry.* 46:8701–8708.
17. Chi, C. N., A. Bach, …, P. Jemth. 2012. Ligand binding by PDZ domains. *Biofactors.* 38:338–348.
18. Doyle, D. A., A. Lee, …, R. MacKinnon. 1996. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell.* 85:1067–1076.
19. Hillier, B. J., K. S. Christopherson, …, W. A. Lim. 1999. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science.* 284:812–815.
20. Songyang, Z., A. S. Fanning, …, L. C. Cantley. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science.* 275:73–77.
21. Stiffler, M. A., J. R. Chen, …, G. MacBeath. 2007. PDZ domain binding selectivity is optimized across the mouse proteome. *Science.* 317:364–369.
22. Steiner, S., and A. Caflisch. 2012. Peptide binding to the PDZ3 domain by conformational selection. *Proteins: Struct., Funct. Bioinf.* 80:2562–2572.
23. Kong, Y., and M. Karplus. 2009. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins: Struct., Funct. Bioinf.* 74:145–154.
24. Mostarda, S., D. Gfeller, and F. Rao. 2012. Beyond the binding site: the role of the β_2-β_3 loop and extra-domain structures in PDZ domains. *PLOS Comput. Biol.* 8:e1002429.
25. Buchli, B., S. A. Waldauer, …, P. Hamm. 2013. Kinetic response of a photoperturbed allosteric protein. *Proc. Natl. Acad. Sci. USA.* 110:11725–11730.
26. Buchenberg, S., V. Knecht, …, G. Stock. 2014. Long-range conformational transition of a photoswitchable allosteric protein: molecular dynamics simulation study. *J. Phys. Chem. B.* 118:13468–13476.
27. Dhulesia, A., J. Gsponer, and M. Vendruscolo. 2008. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J. Am. Chem. Soc.* 130:8931–8939.
28. Shan, Y., E. T. Kim, …, D. E. Shaw. 2011. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133:9181–9183.
29. Dror, R. O., A. C. Pan, …, D. E. Shaw. 2011. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. USA.* 108:13118–13123.
30. Huang, D., and A. Caflisch. 2011. Small molecule binding to proteins: affinity and binding/unbinding dynamics from atomistic simulations. *ChemMedChem.* 6:1578–1580.
31. Gohlke, H., U. Hergert, …, L. Schmitt. 2013. Binding region of alanopine dehydrogenase predicted by unbiased molecular dynamics simulations of ligand diffusion. *J. Chem. Inf. Model.* 53:2493–2498.
32. Buch, I., T. Giorgino, and G. De Fabritiis. 2011. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 108:10184–10189.
33. Magno, A., S. Steiner, and A. Caflisch. 2013. Mechanism and kinetics of acetyl-lysine binding to bromodomains. *J. Chem. Theory Comput.* 9:4225–4232.
34. Gao, X., T. Satoh, …, T. Kataoka. 2001. Identification and characterization of RA-GEF-2, a Rap guanine nucleotide exchange factor that serves as a downstream target of M-Ras. *J. Biol. Chem.* 276:42219–42225.

35. Erdmann, K. S. 2003. The protein tyrosine phosphatase PTP-Basophil/Basophil-like. Interacting proteins and molecular functions. *Eur. J. Biochem.* 270:4789–4798.

36. Abaan, O. D., and J. A. Toretsky. 2008. PTPL1: a large phosphatase with a split personality. *Cancer Metastasis Rev.* 27:205–214.

37. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

38. Van Der Spoel, D., E. Lindahl, …, H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

39. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

40. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

41. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

42. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

43. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

44. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

45. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

46. Blöchliger, N., A. Vitalis, and A. Caflisch. 2013. A scalable algorithm to order and annotate continuous observations reveals the metastable states visited by dynamical systems. *Comput. Phys. Commun.* 184:2446–2453.

47. Blöchliger, N., A. Vitalis, and A. Caflisch. 2014. High-resolution visualization of the states and pathways sampled in molecular dynamics simulations. *Sci. Rep.* 4:6264.

48. Vitalis, A., and A. Caflisch. 2012. Efficient construction of mesostate networks from molecular dynamics trajectories. *J. Chem. Theory Comput.* 8:1108–1120.

49. Zhang, J., P. J. Sapienza, …, A. L. Lee. 2010. Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1E. *Biochemistry.* 49:9280–9291.

50. Dolinsky, T. J., J. E. Nielsen, …, N. A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665–W667.

51. Dolinsky, T. J., P. Czodrowski, …, N. A. Baker. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35:W522–W525.

52. Baker, N. A., D. Sept, …, J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.

53. Krivov, S. V., and M. Karplus. 2006. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B.* 110:12689–12698.

54. Krivov, S. V., and M. Karplus. 2008. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. USA.* 105:13841–13846.

55. Banushkina, P. V., and S. V. Krivov. 2015. Fep1d: a script for the analysis of reaction coordinates. *J. Comput. Chem.* Published online February 25, 2015. http://dx.doi.org/10.1002/jcc.23868.

56. Haq, S. R., C. N. Chi, …, P. Jemth. 2012. Side-chain interactions form late and cooperatively in the binding reaction between disordered peptides and PDZ domains. *J. Am. Chem. Soc.* 134:599–605.

57. Eildal, J. N. N., G. Hultqvist, …, P. Jemth. 2013. Probing the role of backbone hydrogen bonds in protein-peptide interactions by amide-to-ester mutations. *J. Am. Chem. Soc.* 135:12998–13007.

58. Volkov, A. N., J. A. R. Worrall, …, M. Ubbink. 2006. Solution structure and dynamics of the complex between cytochrome *c* and cytochrome *c* peroxidase determined by paramagnetic NMR. *Proc. Natl. Acad. Sci. USA.* 103:18945–18950.

59. Huang, D., and A. Caflisch. 2015. Evolutionary conserved Tyr-169 stabilizes the $\beta 2$-$\alpha 2$ loop of the prion protein. *J. Am. Chem. Soc.* 137:2948–2957.

60. Mahoney, M. W., and W. L. Jorgensen. 2001. Diffusion constant of the TIP5P model of liquid water. *J. Chem. Phys.* 114:363–366.

61. Gianni, S., A. Engström, …, P. Jemth. 2005. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J. Biol. Chem.* 280:34805–34812.

62. Gianni, S., T. Walma, …, G. W. Vuister. 2006. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure.* 14:1801–1809.

63. Gianni, S., S. R. Haq, …, P. Jemth. 2011. Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. *J. Biol. Chem.* 286:27167–27175.

64. Di Silvio, E., D. Bonetti, …, S. Gianni. 2014. The mechanism of binding of the second PDZ domain from the Protein Tyrosine Phosphatase-BL to the Adenomatous Polyposis Coli tumor suppressor. *Protein Eng. Des. Sel.* 27:249–253.

65. Chi, C. N., A. Engström, …, P. Jemth. 2006. Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain. *J. Biol. Chem.* 281:36811–36818.

66. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

67. Sheinerman, F. B., R. Norel, and B. Honig. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10:153–159.

68. Filippakopoulos, P., S. Picaud, …, S. Knapp. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell.* 149:214–231.

69. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.

70. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.

71. Dyson, H. J., and P. E. Wright. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12:54–60.

72. Baker, C. M., and R. B. Best. 2014. Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4:182–198.

73. Chen, T., J. Song, and H. S. Chan. 2014. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.* 30:32–42.

74. Zhang, W., D. Ganguly, and J. Chen. 2012. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. *PLOS Comput. Biol.* 8:e1002353.

75. Rogers, J. M., A. Steward, and J. Clarke. 2013. Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. *J. Am. Chem. Soc.* 135:1415–1422.

76. Ganguly, D., S. Otieno, …, J. Chen. 2012. Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *J. Mol. Biol.* 422:674–684.

77. Ganguly, D., W. Zhang, and J. Chen. 2013. Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins. *PLOS Comput. Biol.* 9:e1003363.

78. De Sancho, D., and R. B. Best. 2012. Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1α with CBP. *Mol. Biosyst.* 8:256–267.
79. Dogan, J., X. Mu, …, P. Jemth. 2013. The transition state structure for coupled binding and folding of disordered protein domains. *Sci. Rep.* 3:2076.
80. Dobson, C. M., A. Šali, and M. Karplus. 1998. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int.Ed. Engl.* 37:868–893.
81. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.
82. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
83. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
84. Best, R. B., G. Hummer, and W. A. Eaton. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA.* 110:17874–17879.
85. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

Supporting Material to
Peptide Binding to a PDZ Domain by Electrostatic Steering via Non-Native Salt Bridges

Nicolas Blöchliger, Min Xu, and Amedeo Caflisch
Department of Biochemistry
University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich

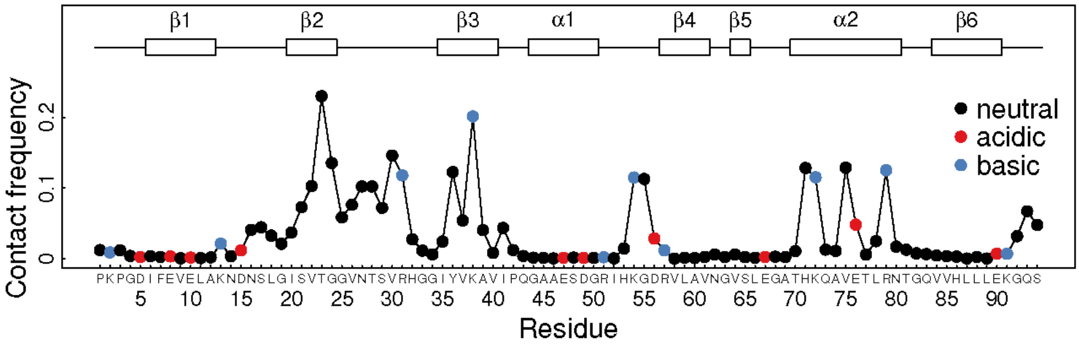


Fig. S1 Contact frequencies between protein residues and the peptide. For every residue, the average contact frequency over the peptide residues is plotted. Secondary structure elements of the protein are indicated on top. See Materials and Methods for details.

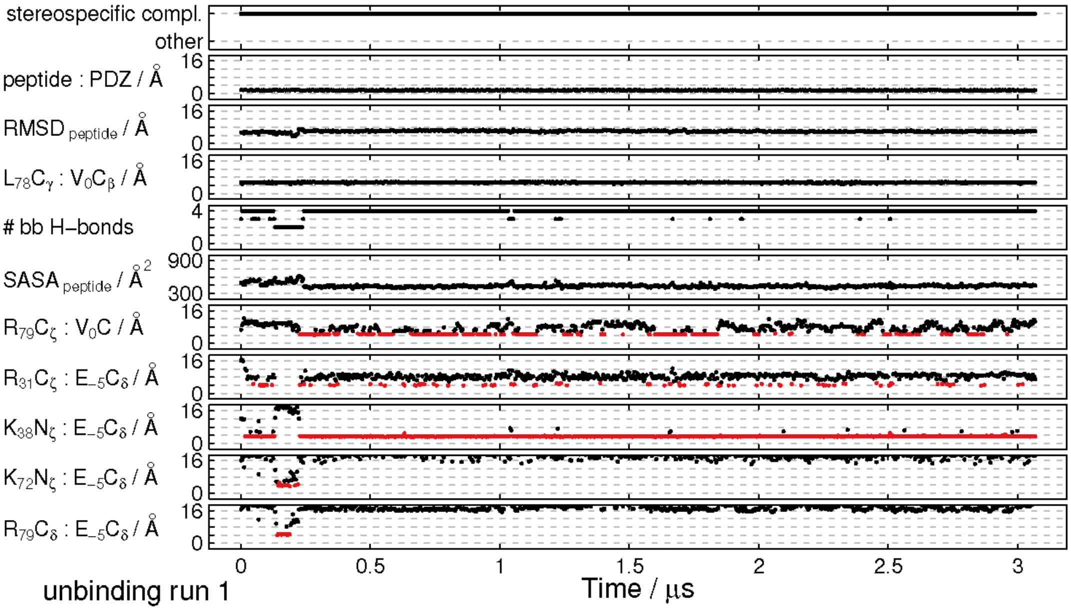


Fig. S2 Analysis of the first unbinding run. Distance between peptide and PDZ domain (peptide : PDZ), RMSD of the peptide C$_\alpha$ atoms after alignment on the PDZ domain, number of backbone hydrogen bonds formed (i.e. distances between the carbonyl oxygen of Ile20 and the NH group of Val0, between both polar groups of Val22 and Ser-2, and between the NH group of Gly24 and the carbonyl oxygen of Gln-4 below 3.5 Å), solvent accessible surface area of the peptide (SASA$_{\text{peptide}}$) and distances between selected atom pairs are shown. The distance between the Leu78 C$_\gamma$ atom and the Val0 C$_\beta$ atom reports on burial of the Val0 side chain in the binding pocket. Distances indicating salt bridge formation are coloured red if their value is below 5.25 Å. All quantities are shown as median values in a window of 1 ns. The top panel indicates the snapshots we classified as representing the stereospecific complex. Corresponding figures are given in Figs. S3–S11 for the other unbinding runs and in Figs. S12–S21 for the binding runs, respectively. Note that different distances are shown in Figs. S3–S21 to illustrate salt bridges.

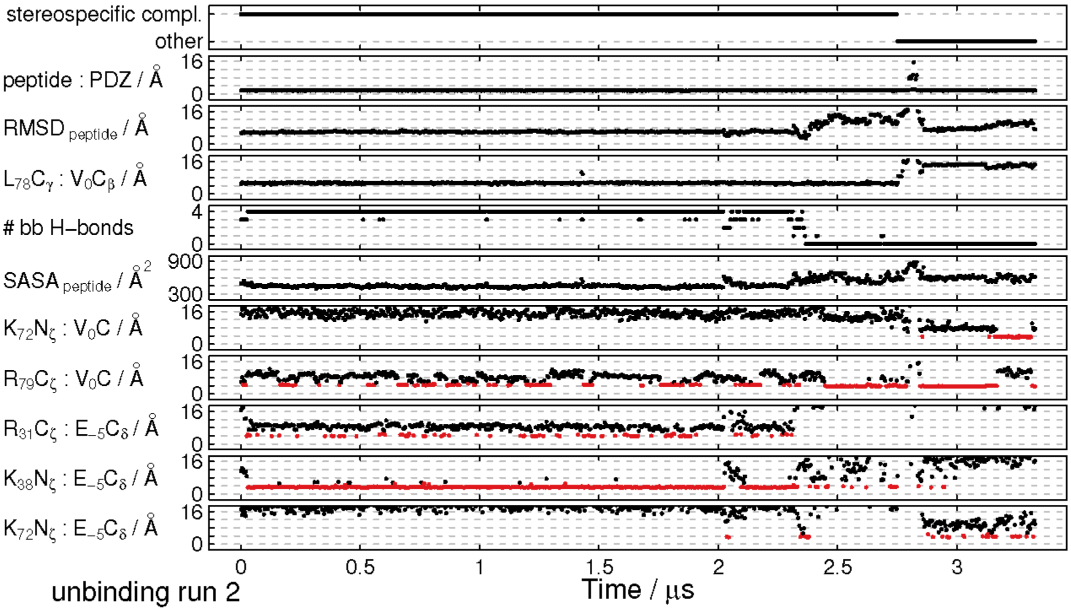


Fig. S3 Analysis of the second unbinding run. Similar to Fig. S2.

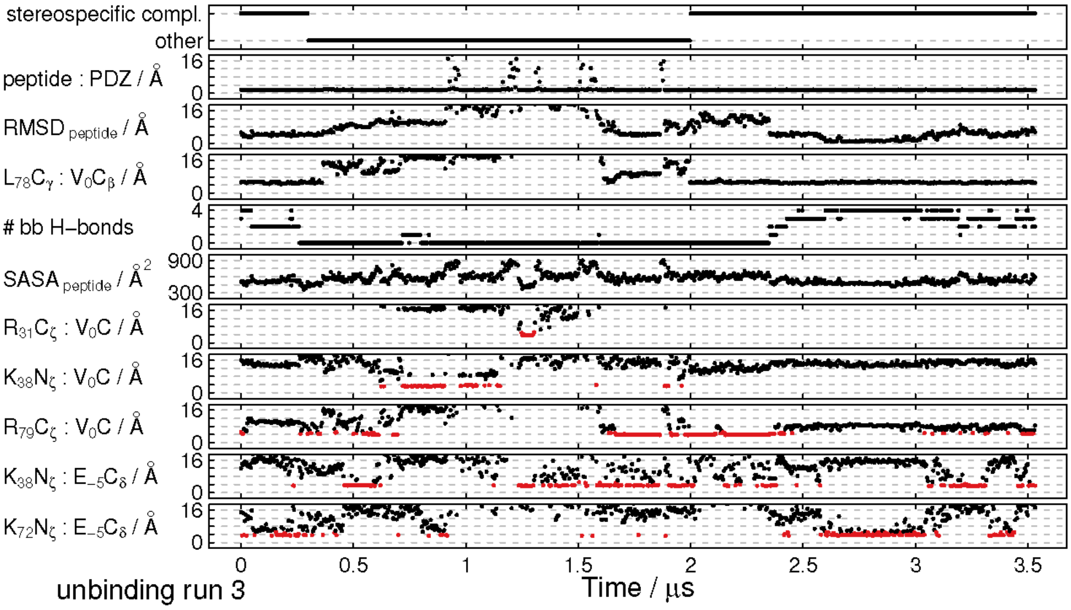


Fig. S4 Analysis of the third unbinding run. Similar to Fig. S2.

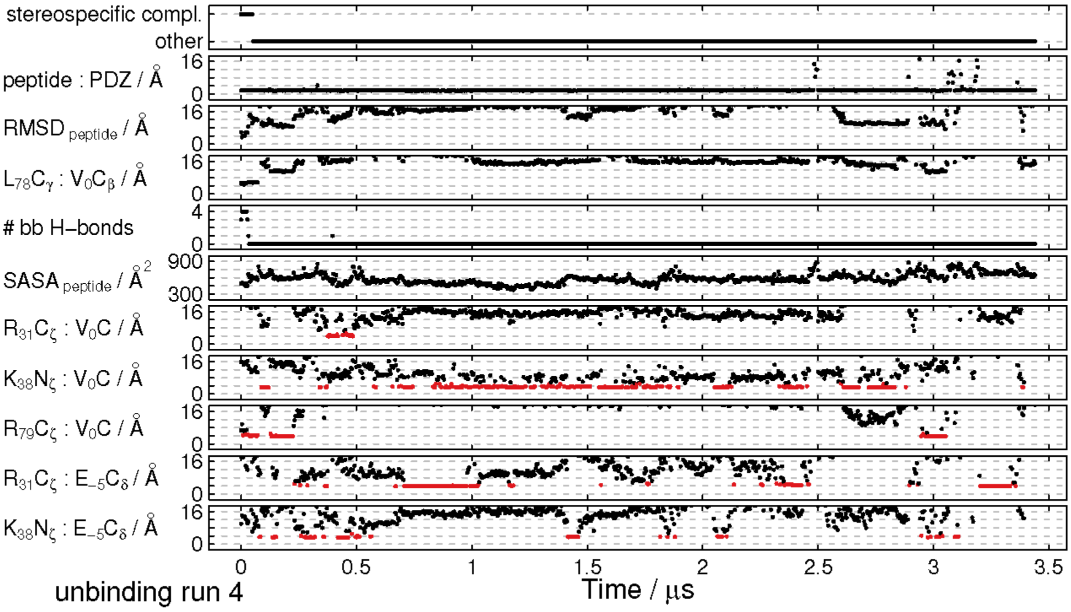


Fig. S5 Analysis of the fourth unbinding run. Similar to Fig. S2.

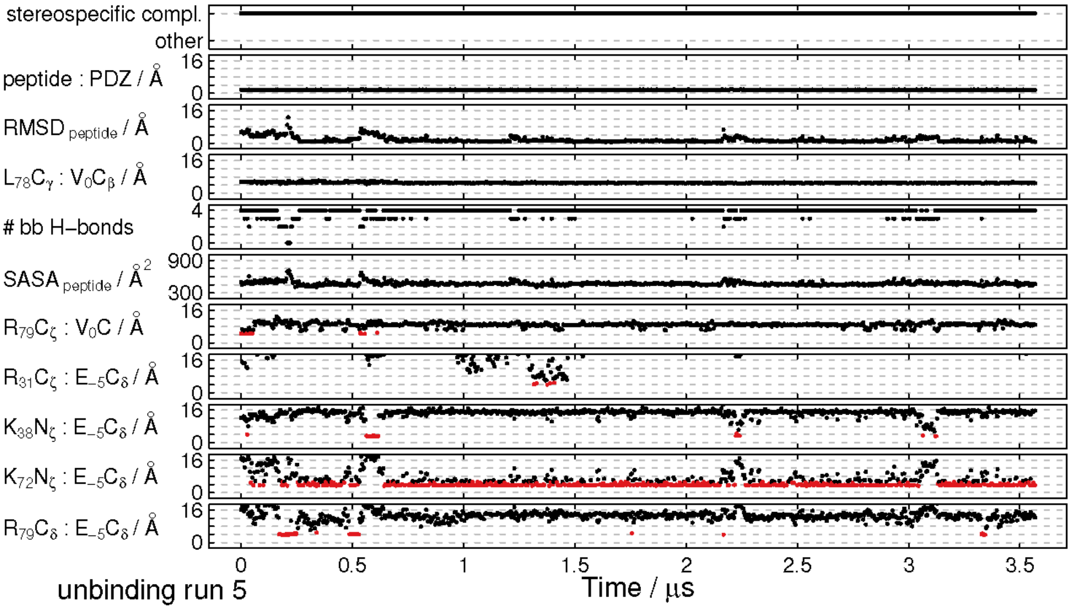


Fig. S6 Analysis of the fifth unbinding run. Similar to Fig. S2.

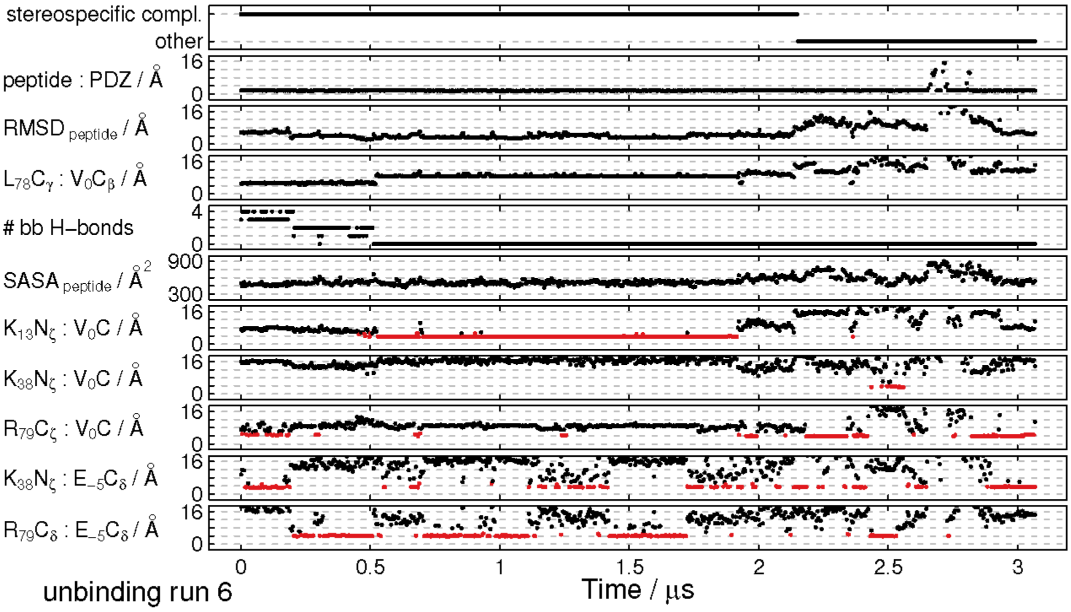


Fig. S7 Analysis of the sixth unbinding run. Similar to Fig. S2.

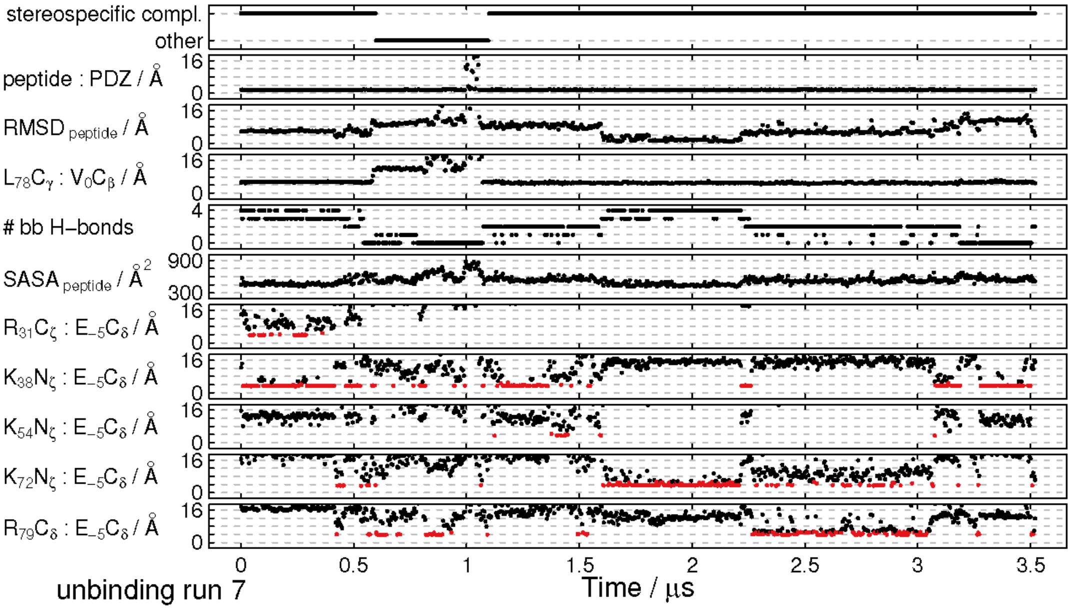


Fig. S8 Analysis of the seventh unbinding run. Similar to Fig. S2.

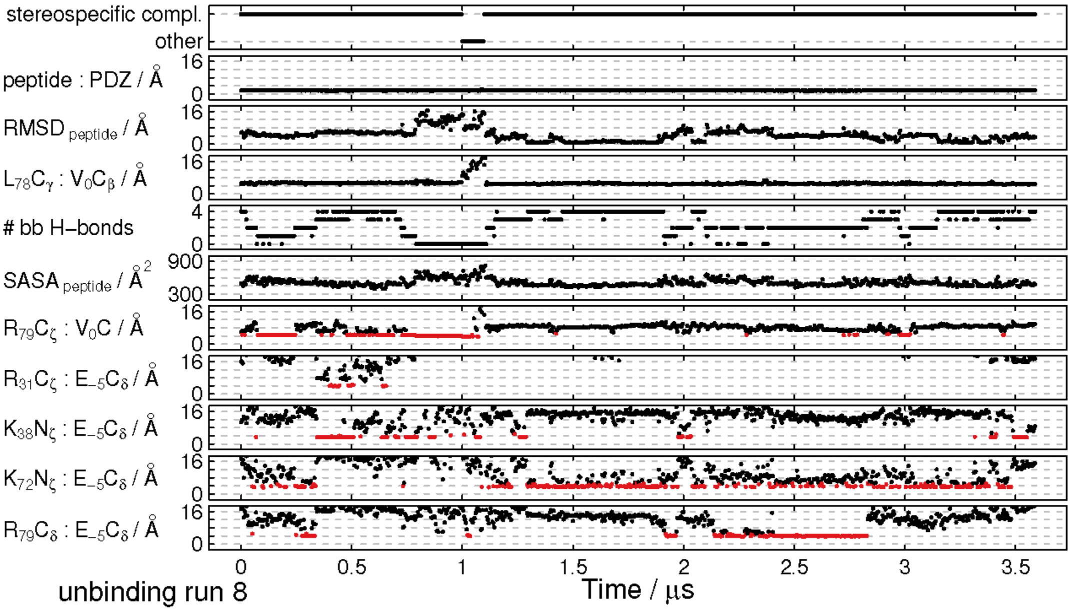


Fig. S9 Analysis of the eighth unbinding run. Similar to Fig. S2.

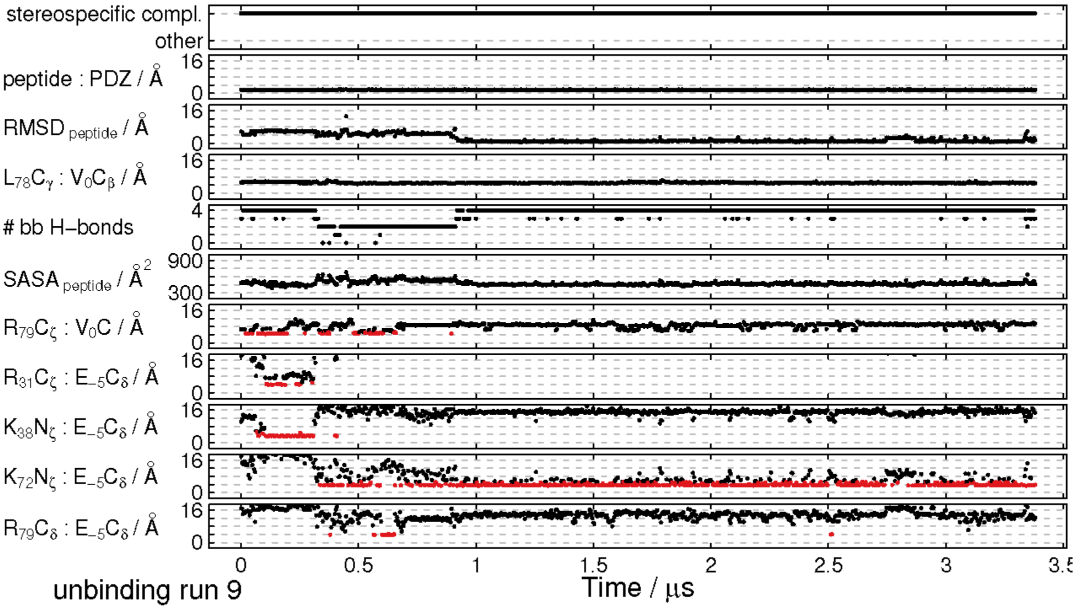


Fig. S10 Analysis of the ninth unbinding run. Similar to Fig. S2.

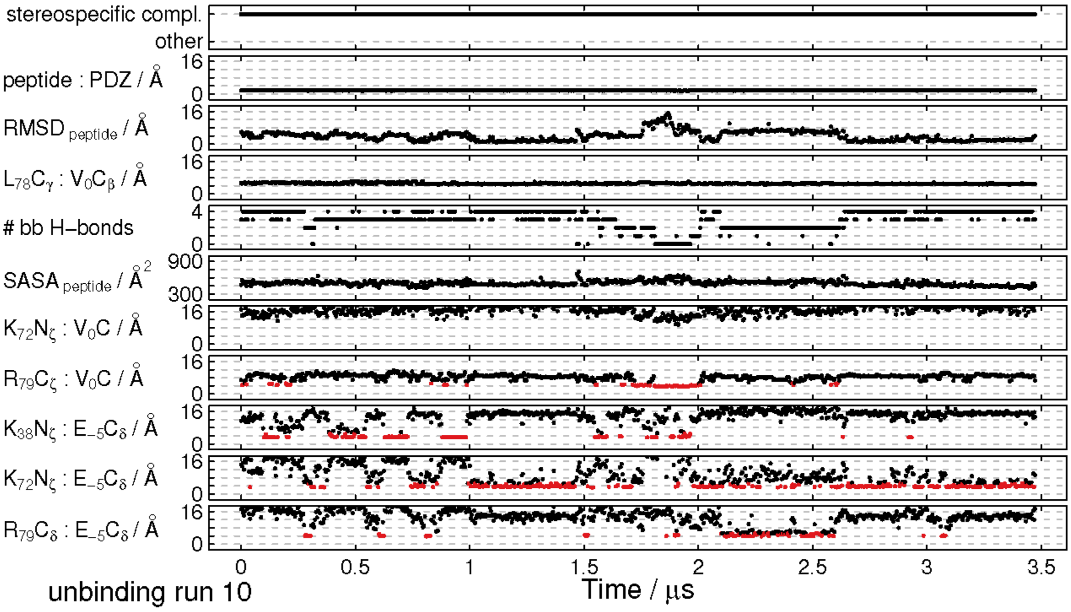


Fig. S11 Analysis of the tenth unbinding run. Similar to Fig. S2.

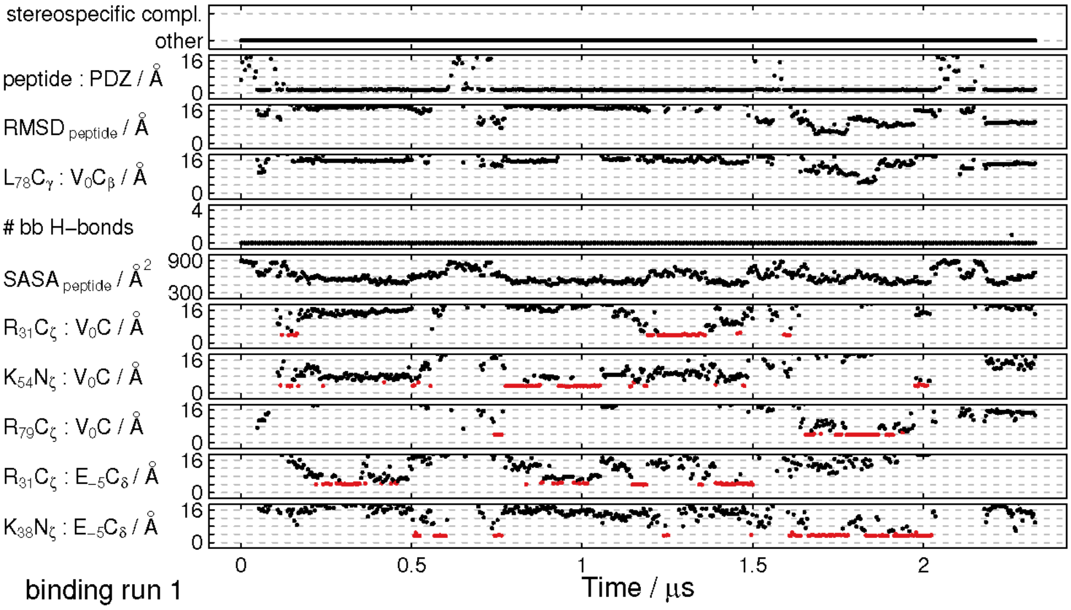


Fig. S12 Analysis of the first binding run. Similar to Fig. S2.

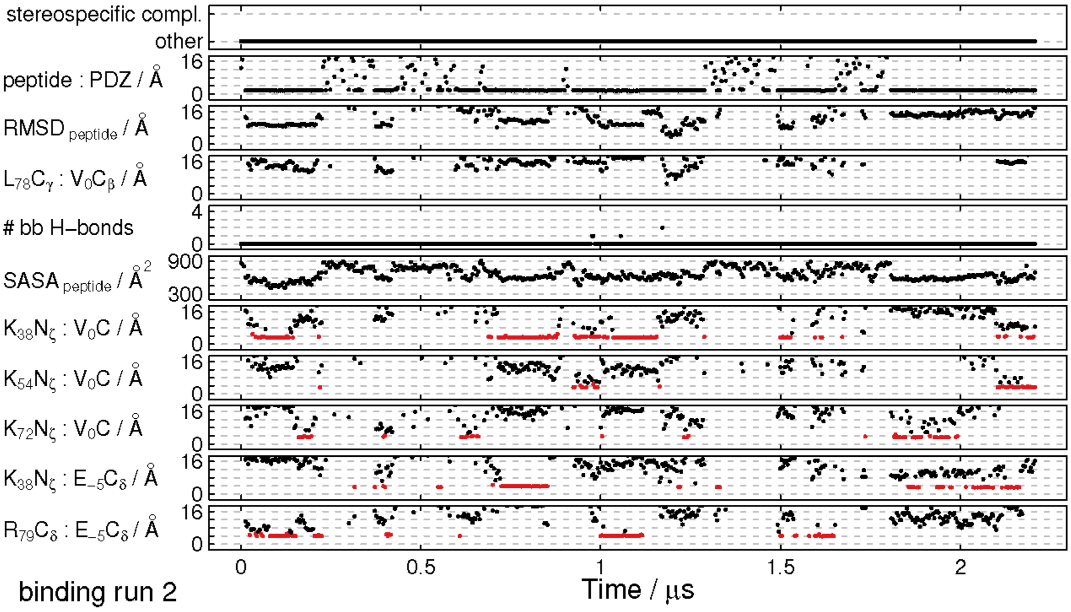


Fig. S13 Analysis of the second binding run. Similar to Fig. S2.

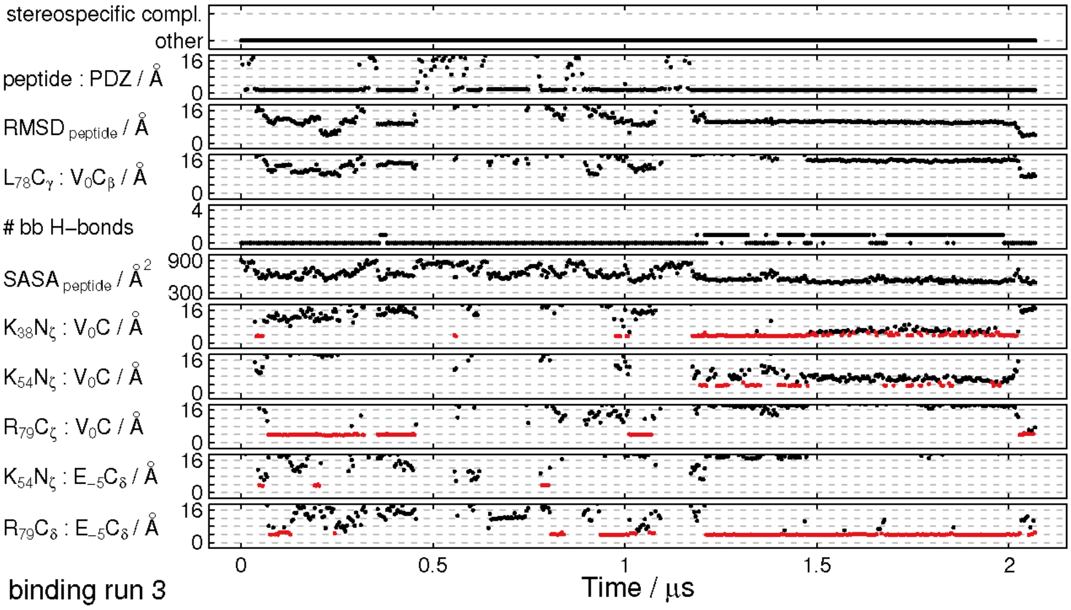


Fig. S14 Analysis of the third binding run. Similar to Fig. S2.

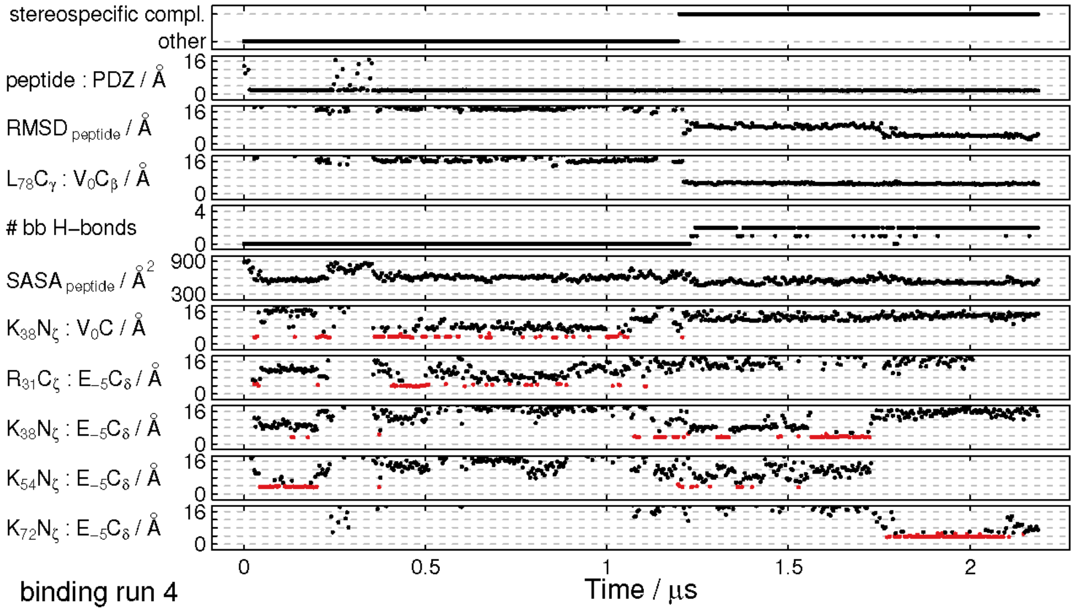


Fig. S15 Analysis of the fourth binding run. Similar to Fig. S2.

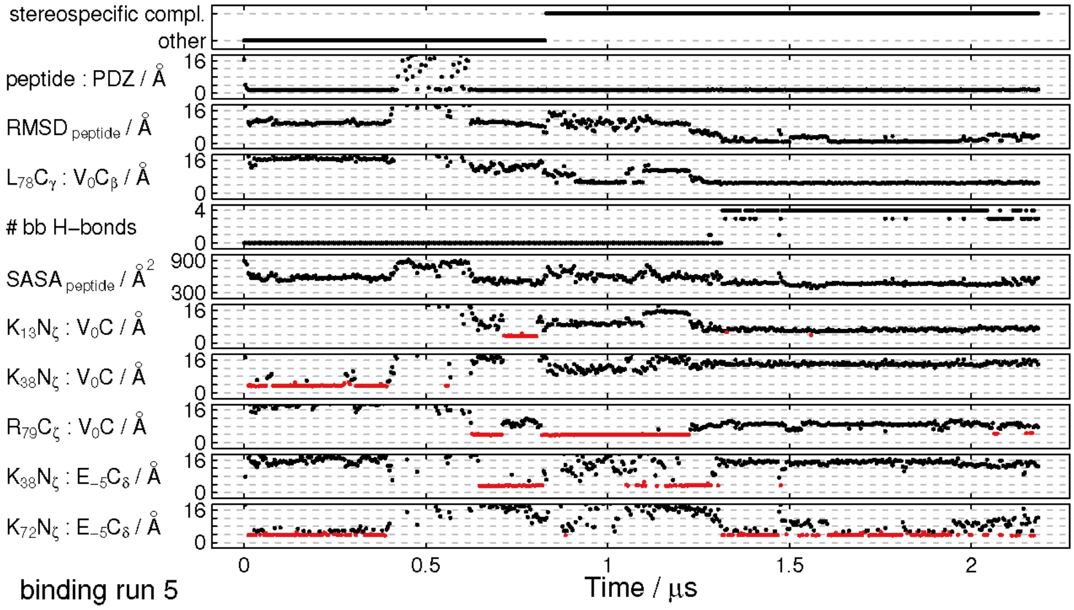


Fig. S16 Analysis of the fifth binding run. Similar to Fig. S2.

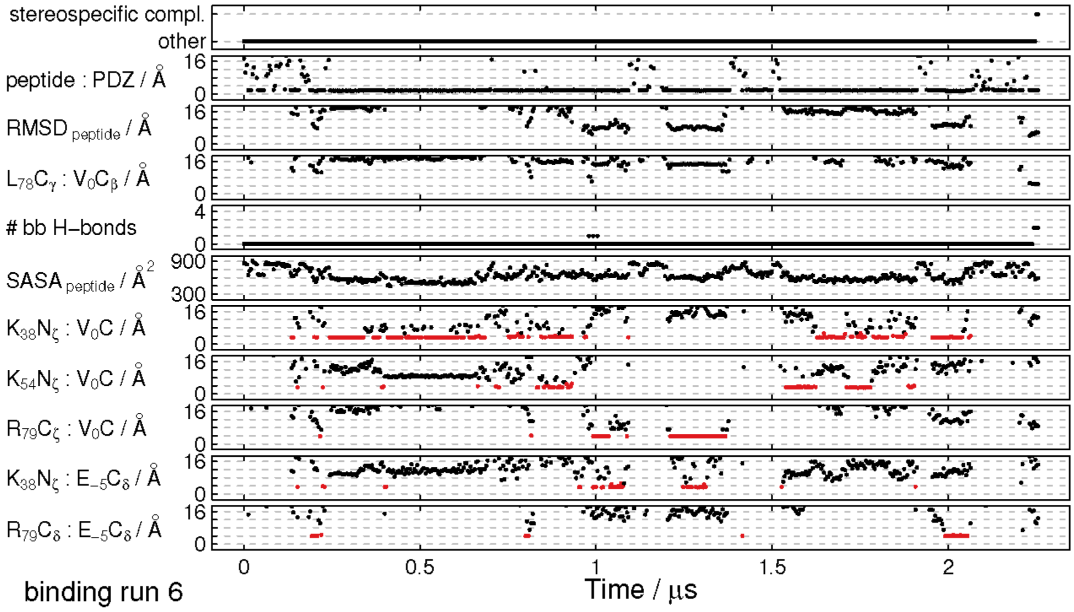


Fig. S17 Analysis of the sixth binding run. Similar to Fig. S2.

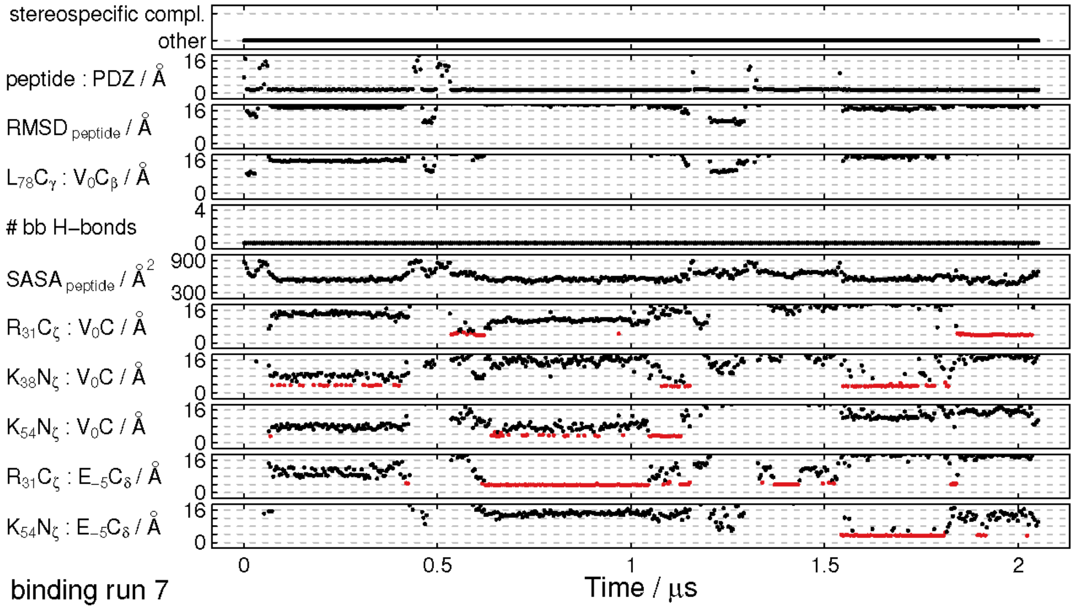


Fig. S18 Analysis of the seventh binding run. Similar to Fig. S2.

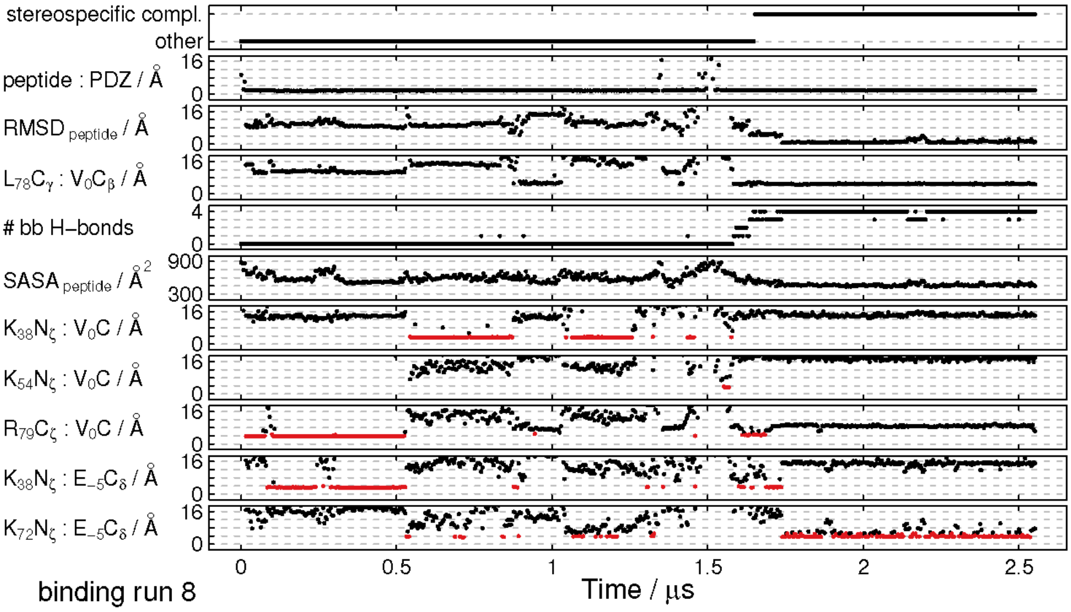


Fig. S19 Analysis of the eighth binding run. Similar to Fig. S2.

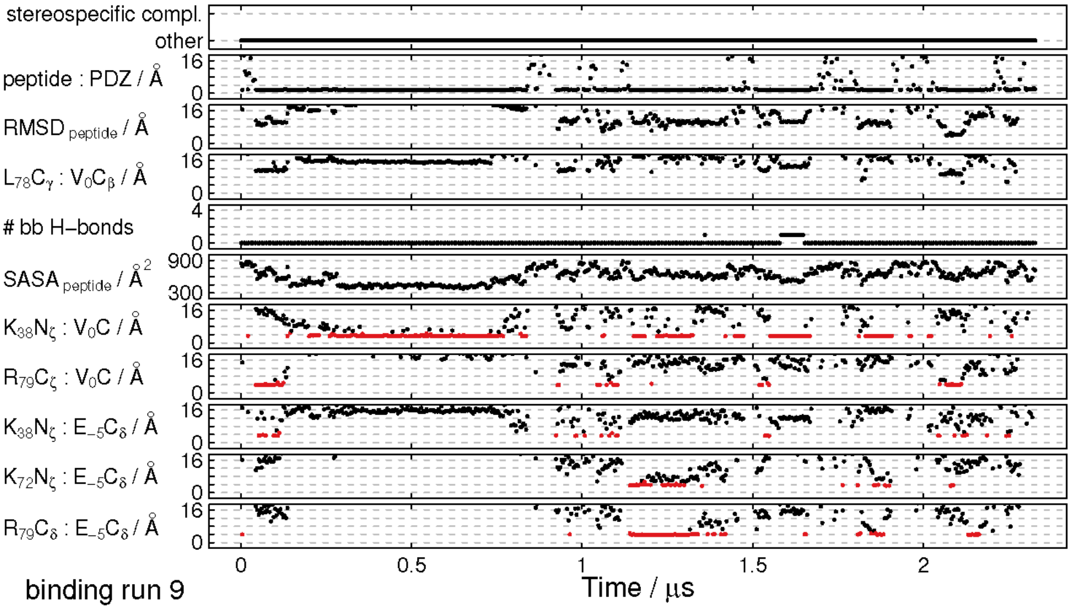


Fig. S20 Analysis of the ninth binding run. Similar to Fig. S2.

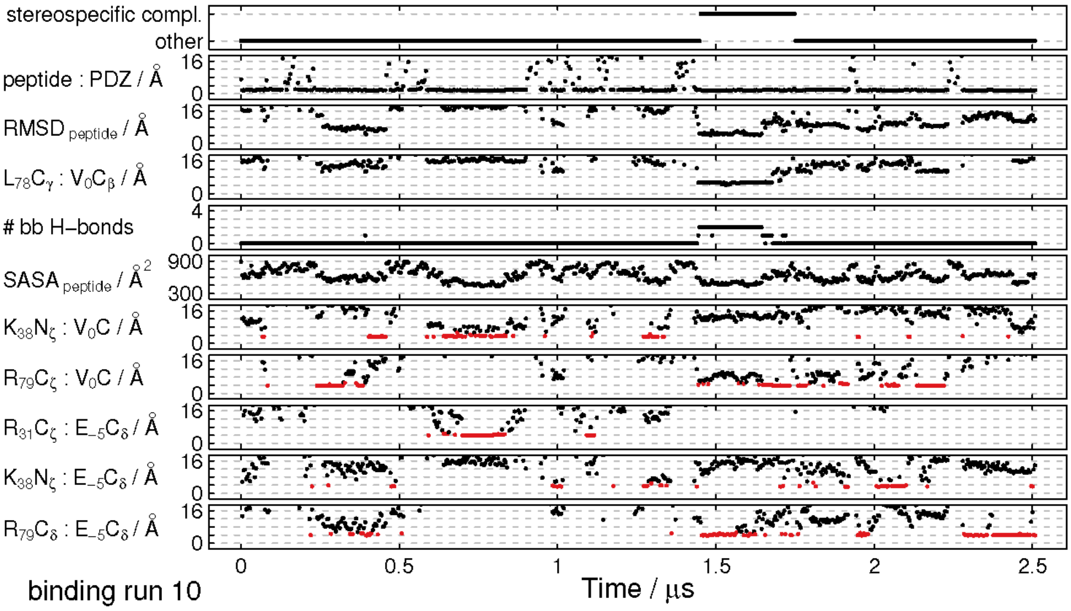


Fig. S21 Analysis of the tenth binding run. Similar to Fig. S2.

Group	Atom pair			
carboxylate-binding loop	Ser17	H_γ	Val0	C
	Leu18	H_N	Val0	C
	Gly19	H_N	Val0	C
	Ile20	H_N	Val0	C
$\beta2$	Ile20	O	Val0	H_N
	Val22	H_N	Ser-2	O
	Val22	O	Ser-2	H_N
	Gly24	H_N	Gln-4	O
$\alpha2$	His71	N_ϵ	Ser-2	H_γ
	Arg79	C_ζ	Ala-1	O
	Arg79	C_ζ	Val0	C
hydrophobic pocket	Leu18	C_β	Val0	C_β
	Leu18	C_γ	Val0	C_β
	Ile20	$C_{\gamma,1}$	Val0	C_β
	Ile20	C_δ	Val0	C_β
	Val22	C_β	Val0	C_β
	Val75	C_β	Val0	C_β
	Leu78	C_β	Val0	C_β
	Leu78	C_γ	Val0	C_β
	Arg79	C_γ	Val0	C_β
N-terminus of ligand	Thr23	C_α	Glu-5	C_α
	Thr23	C_α	Gln-4	C_α
	Thr23	C_α	Gln-4	C_δ
	Thr23	C_α	Val-3	C_α
	His71	C_α	Glu-5	C_α
	His71	C_α	Gln-4	C_α
	His71	C_α	Gln-4	C_δ
	His71	C_α	Val-3	C_α
	Asn27	N_δ	Gln-4	O_ϵ

Table S1 Distance function used for SAPPHIRE plot. This table lists the 29 atom pairs for the distance function used for the SAPPHIRE plot shown in Fig. 2 of the main text. Distances are grouped structurally. The group 'hydrophobic pocket' contains atoms from the hydrophobic pocket surrounding the side chain of Val0 in the crystal structure, and the group 'N-terminus of ligand' is meant to capture the orientation of the N-terminal part of the ligand with respect to the protein. The distance between two snapshots i and j is given by $\sqrt{\sum_{k=1}^{29}(d_k^i - d_k^j)^2}$ where d_k^i is the distance between the k-th atom pair in snapshot i.

Chapter 4

Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics

Xu, M., Unzue, A., Dong, J., Spiliotopoulos, D., Nevado, C., Caflisch, A.,

Journal of Medicinal Chemistry 2015.

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics

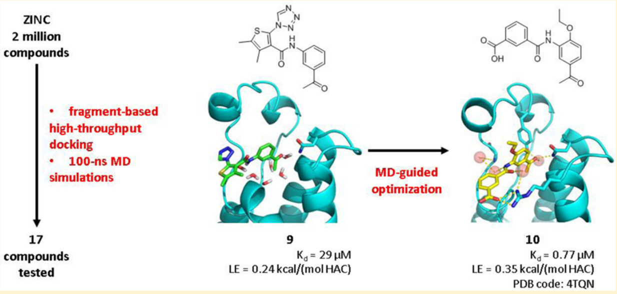
Min Xu,[†] Andrea Unzue,[‡] Jing Dong,[†] Dimitrios Spiliotopoulos,[†] Cristina Nevado,*[,‡] and Amedeo Caflisch*[,†]

[†]Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland

[‡]Department of Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland

Ⓢ *Supporting Information*

ABSTRACT: We have identified two chemotypes of CREBBP bromodomain ligands by fragment-based high-throughput docking. Only 17 molecules from the original library of two-million compounds were tested in vitro. Optimization of the two low-micromolar hits, the 4-acylpyrrole **1** and acylbenzene **9**, was driven by molecular dynamics results which suggested improvement of the polar interactions with the Arg1173 side chain at the rim of the binding site. The synthesis of only two derivatives of **1** yielded the 4-acylpyrrole **6** which shows a single-digit micromolar affinity for the CREBBP bromodomain and a ligand efficiency of 0.34 kcal/mol per non-hydrogen atom. Optimization of the acylbenzene hit **9** resulted in a series of derivatives with nanomolar potencies, good ligand efficiency and selectivity (see Unzue, A.; Xu, M.; Dong, J.; Wiedmer, L.; Spiliotopoulos, D.; Caflisch, A.; Nevado, C. Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b00172). The in silico predicted binding mode of the acylbenzene derivative **10** was validated by solving the structure of the complex with the CREBBP bromodomain.



■ INTRODUCTION

Bromodomains are α-helical modules of about 110 residues that bind the acetylated side chain of lysine most notably (but not exclusively) in histone proteins.[1] Bromodomain-containing proteins have up to six of these modules. Some of these proteins play a crucial role in numerous cellular processes, such as gene regulation.[2] The three-dimensional structures of most of the 61 human bromodomains have been solved by X-ray crystallography[3] and in a few cases by NMR spectroscopy.[4]

Small molecules inhibiting CREBBP bromodomain acetyllysine-binding activity have been identified in the last years (Figure 1). Zhou et al. used NMR spectroscopy (the ^{1}H−^{15}N HSQC spectra) to screen a small compound library and were the first to identify two compounds that block the interaction between p53 and KAc382-CREBBP at 100 and 50 μM, respectively (compounds **A** and **B** in Figure 1).[4a] The same group identified ischemin, which has an IC_{50} of 5 μM for the CREBBP bromodomain (**C**, Figure 1),[4b] and synthesized a p53-like cyclic peptide that binds the CREBBP bromodomain with an IC_{50} of 5 μM.[5] The identification of *N*-methyl-2-pyrrolidone as a weak antagonist of acetylated peptide-binding by the CREBBP bromodomain (**D**, Figure 1)[6] ignited studies on methyl-bearing heterocycles by Hewings and co-workers, who developed substituted 3,5-dimethylisoxazoles compounds targeting BET bromodomains, and a poorly selective CREBBP bromodomain inhibitor (**E**, Figure 1).[7] In a study that focused on 40 crystal structures of the N-terminal bromodomain of BRD2 in the complex with rigid fragments, Chung et al. disclosed also the crystal structure of the complex of the CREBBP bromodomain and paracetamol (**F**, Figure 1).[8] Hay and collaborators described some nonselective CREBBP ligands while aiming to identify BET inhibitors (**G**, Figure 1).[9] Another 3,5-dimethylisoxazole derivative potently and selectively binding the CREBBP bromodomain (called SGC-CBP30) has been reported recently (**H**, Figure 1).[10] The SGC has also reported the development of a nanomolar CREBBP bromodomain inhibitor, I-CBP112, which exhibits weak binding activity toward the BET bromodomains (**I**, Figure 1).[11] Rooney and co-workers developed a dihydroquinoxalinone series ultimately leading to a modestly selective nanomolar CREBBP bromodomain inhibitor (**J**, Figure 1).[12] A series of [1,2,4]triazolo[4,3-*a*]phthalazines was shown recently to bind unselectively to diverse bromodomains including CREBBP (**K**, Figure 1).[13]

To identify small molecules that bind selectively to the CREBBP bromodomain, we have performed in silico screening with two crystal structures of this target. ALTA (anchor-based library tailoring),[14] a four-step procedure developed in house for docking large libraries of compounds, was used for virtual

Special Issue: Epigenetics

Received: January 29, 2015

DOI: 10.1021/acs.jmedchem.5b00171
J. Med. Chem. XXXX, XXX, XXX−XXX

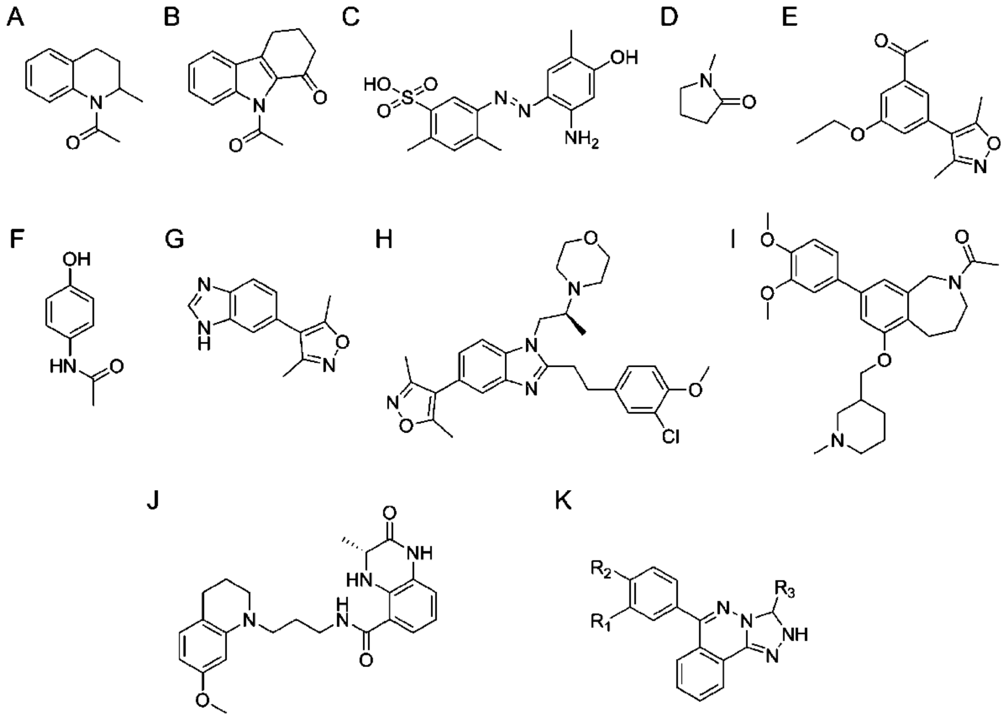


Figure 1. Known CREBBP inhibitors reported in the literature. Compounds **B**, **C**, **D**, **E**, **H**, **I**, and **J** can be found in complex with the CREBBP bromodomain in the PDB files 2D82, 2L84, 3P1D, 3SVH, 4NR7, 4NR6 and 4NYX, respectively.

screening. In the first step, the molecules are decomposed into rigid fragments by an algorithm that cuts at rotatable bonds.[15] In the second step, the fragments are docked and ranked using an energy evaluation that takes into account solvation effects in the continuum dielectric approximation.[16] Third, the procedure identifies the molecules that contain at least one of the top ranking fragments. The final step consists of the flexible docking of the retrieved molecules followed by energy minimization and evaluation of desolvation penalties.

Here we report on the in silico discovery by the ALTA procedure of two hits (compounds **1** and **9**) for the CREBBP bromodomain. To the best of our knowledge, this is the first in silico screening that targets the CREBBP bromodomain. Multiple molecular dynamics (MD) runs with explicit solvent were carried out to further validate the binding mode suggested by docking and guide chemical synthesis for hit optimization. The synthesis of only two derivatives of the 4-acylpyrrole hit **1** resulted in a single-digit micromolar ligand of the CREBBP bromodomain with excellent ligand efficiency. The optimization of the acylbenzene hit **9** is described in detail in ref 17.

■ DOCKING RESULTS

Docking of Fragments. The ALTA procedure started with the decomposition of the nearly two million compounds in the ZINC leads-now library (version of October 2012) by the DAIM program[15] (Figure 2). To target the conserved Asn involved in binding the natural ligand acetyllysine (numbered 1168 in the PDB structures 3P1C and 3SVH), only fragments with a hydrogen bond donor or acceptor were kept. Although the natural ligand (acetyllysine) acts as hydrogen bond acceptor for the amide side chain of the conserved Asn, it was decided to also consider fragments with a hydrogen bond donor because our previous MD simulations revealed that the amide of the Asn side chain occupies both possible orientations (Figure 2 of ref 18). Further removal of fragments without a ring yielded a set of 106 534 fragments.

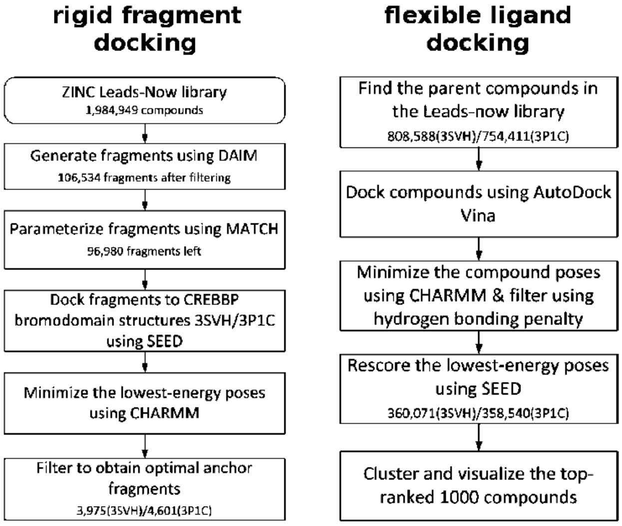


Figure 2. (Left) Flowchart of fragment docking which is the initial phase of the ALTA (anchor-based library tailoring) procedure.[14a] The program DAIM automatically decomposes molecules into fragments by cutting at rotatable bonds.[15] The fragment-docking tool SEED[16] required about 5 s per fragment on a single core of an i7 CPU at 2.8 GHz. CHARMM[20] minimization of the fragment with rigid CREBBP took about 2 s per fragment. (Right) Flowchart of flexible ligand docking which is the second phase of the ALTA procedure. Flexible ligand docking by AutoDock Vina required about 1 min per compound.

The X-ray structures of the CREBBP bromodomain in complex with the natural ligand acetyllysine (PDB code 3P1C) and the synthetic ligand 3,5-dimethylisoxazole (3SVH) show different orientations of the side chain of the so-called gatekeeper residue, Val1174 (Figure 3).[19] Furthermore, the solvent-exposed side chain of Arg1173 seems very flexible in the two crystallographic structures. In 3P1C, it is just above the Leu-Pro-Phe segment (called LPF, in the following; this tripeptide stretch corresponds to the WPF segment in most bromodomains). On the contrary, in 3SVH the Arg1173 guanidinium points outward (Figure 3). Given these differences, it was decided to use both structures for docking. After the preparation of the two protein structures (see Experimental Section), the fragment library was docked into the binding pocket by the program SEED.[16]

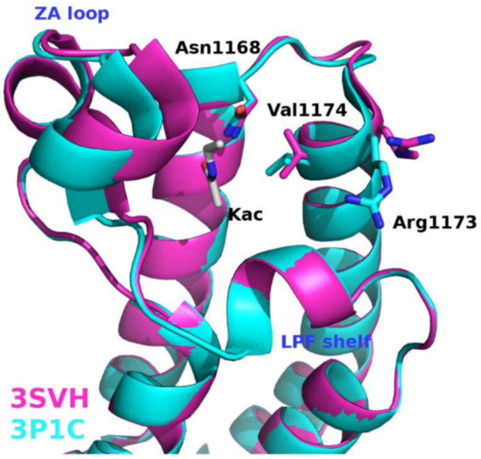


Figure 3. Overlap of the two crystal structures of the CREBBP bromodomain used for docking. The helical secondary structure is shown by ribbons. The conserved Asn1168 and the residues with variable side chain orientation that are mentioned in the text are shown in sticks. The physiological ligand (acetylated lysine) is shown as sticks (carbon atoms in gray).

Two additional filters were applied to reduce the number of docked fragments to about 4000 for each of the two CREBBP structures (Figure 2, left): (a) SEED total energy efficiency more favorable than −0.125 kcal mol^{-1} g^{-1} and (b) hydrogen bonding penalty of less than or equal to 1.[21] These criteria efficiently enriched the fragments involved in hydrogen bond(s) with the Asn1168 side chain, a key interaction present in the complex of the natural ligand with CREBBP (Figure 3). Quantitatively, after filtering there were 91% and 82% of fragments forming at least one hydrogen bond with the Asn1168 side chain for 3SVH and 3P1C, respectively, while before filtering there were 33% and 41%, respectively.

To assess the predictive ability of the procedure used for fragment ranking, six fragments from known inhibitors were also docked by SEED (Figure 4). Importantly, the pose of these fragments predicted by SEED matches within 1 Å the binding mode observed in the crystal structures.

For each of the two crystal structures used for docking, the receiver-operator characteristic (ROC) plot generated using about 100 000 fragments with optimal SEED energy efficiency (considered as false positives) and six fragments from known inhibitors (true positives) gives an area under the curve of 0.985. The individual ranks of the six known fragments are listed in Figure 4. Thus, the very large area under the curve and the good ranks of the fragments from known inhibitors provide evidence that the filtering and ranking procedure is able to prioritize true positives.

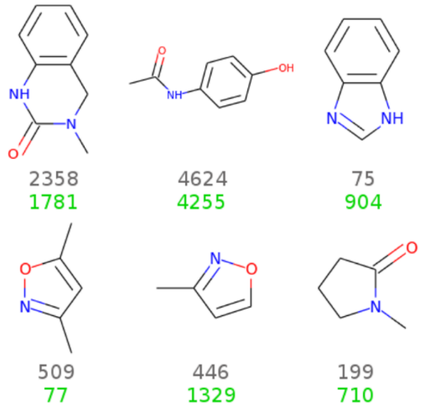


Figure 4. Validation of the filtering and ranking procedures. The figure shows the six fragments from known CREBBP and BET bromodomain inhibitors[22] that were used as validation set, i.e., considered as true positives. The two integers below each fragment are their ranks among the 96 724 putatively inactive fragments docked into the 31PC structure (gray) and the 100 027 putatively inactive fragments docked into 3SVH (green), respectively.

Docking of Compounds with Optimal Anchor Fragments. About 4000 fragments were used to retrieve the molecules in the ZINC library which they originated from (called parent compounds hereafter). There were 808 588 (3SVH)/754 411 (3P1C) parent compounds in total and for each of them at most 15 poses were generated by AutoDock Vina.[23] All the poses were then minimized, filtered, and rescored using the SEED total energy (Figure 2, right). In the end, the top-ranking poses of 1000 compounds according to the SEED total energy were chosen and analyzed. They were clustered based on the functional groups interacting with the conserved Asn1168, which in the following is referred to as the "head group". The chemical space of the head groups was rather condensed (Figure 5). Thus, only 20 compounds (representing all of the head groups) were chosen for further investigation by MD simulations to verify the stability of the binding mode.

This analysis ruled out three compounds, as they moved out of the binding site within 100 ns (Figure S1 in Supporting Information). The remaining 17 molecules were purchased and tested by a competition binding assay (BROMOScan at DiscoveRx).[24] Two of these compounds (**1** and **9**), both featuring an acetylated aromatic ring, showed an equilibrium dissociation constant (K_d) in the low μM range (Figure 6, Figures S2 and S5). Note that the acylaryl moieties of compounds **1** and **9** do not appear in Figure 5 because they were not among the most frequent in the 1000 top ranked compounds. Nonetheless, the aforementioned clustering and MD simulations were able to identify them.

■ EXPERIMENTAL RESULTS

Optimization of 4-Acylpyrrole Hit 1 Guided by MD Simulation Results. The binding mode of compound **1** obtained by rigid-protein docking into CREBBP bromodomain indicates that its acyl group acts as hydrogen bond acceptor for the side chain of the conserved Asn1168 (Figure S2A). Thus, five additional commercially available 4-acylpyrrole derivatives

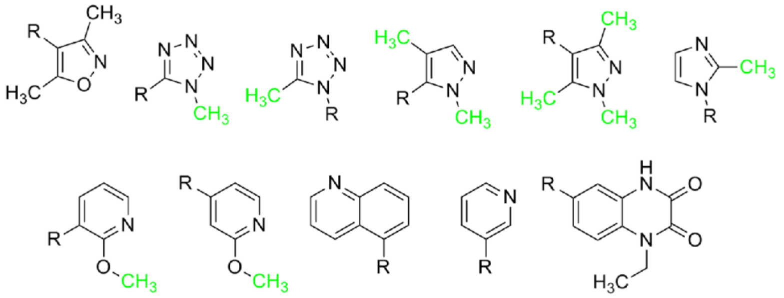
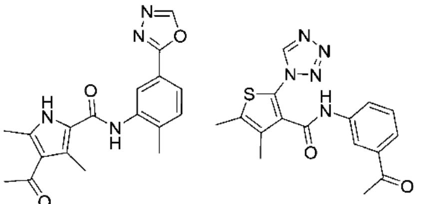


Figure 5. Most frequent anchor fragments in the 1000 top ranked compounds. The methyl groups in green are H atoms in some of the compounds.



1
K_d = 13 μM

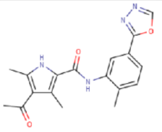
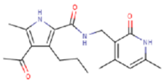
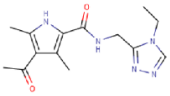
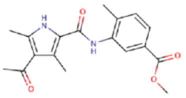
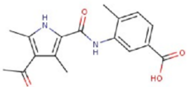
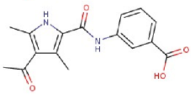
9
K_d = 29 μM

Figure 6. CREBBP bromodomain hits identified by the fragment-based ALTA procedure. In the binding mode predicted by docking the acetyl group accepts a hydrogen bond from the side chain of Asn1168. K_d values for the CREBBP bromodomain were measured by the competition binding assay[24] (Figure S5).

were tested. Among these, compounds **2**, **3**, and **4** showed activity at 50 μM in a competition binding assay (Table 1).[24]

Next, we analyzed the predicted binding modes of compounds **1**–**4**, which share the presence of an amide as a linker connecting the 4-acylpyrrole (the compounds' "head group") with a partially exposed moiety (called "tail group" henceforth). Their binding poses, as predicted by docking, are overall similar (Figure S3). The acyl oxygen is involved as acceptor in a hydrogen bond with the amide side chain of Asn1168. The two alkyl substituents on the pyrrole ring further strengthen the binding mode by van der Waals interactions with the side chains of Leu1120 on the ZA loop and Val1174. The linker amide establishes a hydrogen bond with one of the structured water molecules while maintaining the conserved water molecule network within the binding site. Interestingly,

Table 1. Experimental Validation of 4-Acylpyrroles Binding to the CREBBP Bromodomain[e]

Compound	2D structure	Thermal shift (°C)[a]	K_d [%Ctrl @50 μM][b]	IC_{50}[c]	Ligand efficiency[d]
1		0.3	13 μM [9%]	> 100 μM	0.27
2		0.3	[3%]	ND	ND
3		1.3	[5%]	ND	ND
4		ND	[20%]	ND	ND
5		1.4	ND	> 10 μM	ND
6		2.3	4.2 μM	1.5 μM	0.34, 0.37

[a]The thermal shift experiments were carried out at a protein concentration of 2 μM and a ligand concentration of 100 μM. At least 16 measurements were acquired for each compound. All standard error values were below 0.2 °C. [b]Dissociation constants were measured by a competition binding assay (Figure S5).[24] The percentage of the measured signal at an inhibitor concentration of 50 μM with respect to the negative control DMSO is given in brackets with lower values indicating stronger binding. Measurements at each dose were done in duplicate. [c]The IC_{50} values were determined by a TR-FRET assay (see Supporting Information). The values are averages of duplicate dose−response measurements. Compound **1** is a false negative in the TR-FRET assay. [d]The ligand efficiency has units of kcal/(mol HAC) where HAC is the heavy atom count. [e]Hit **1** was identified by the ALTA procedure. Compounds **2**−**4** were identified by substructure search, and compounds **5** and **6** were synthesized (see Scheme 1). ND: not determined.

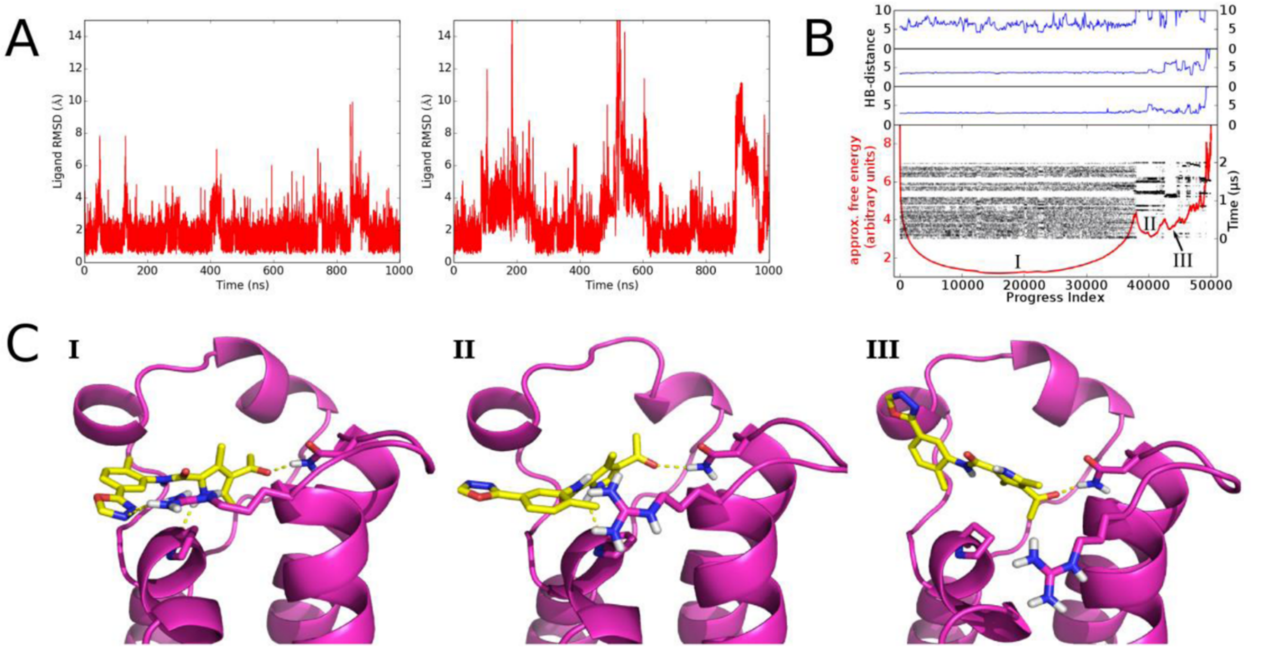


Figure 7. Structural stability and multiple binding modes of compound **1** in complex with the CREBBP bromodomain. (A) Two independent MD runs were started from the binding mode obtained by docking using different initial velocities. The 100 000 snapshots (saved every 20 ps) were overlapped using the Cα atoms of the helical segments of the CREBBP bromodomain, and the time series of the RMSD of the heavy atoms of compound **1** was calculated. (B) MD simulations suggest multiple binding modes for compound **1**. The free-energy profile[27] of the binding of compound **1** to the CREBBP bromodomain (red curve) is shown together with the trace of the temporal evolution of the system during the 2 μs MD sampling (black dots). The dynamical trace illustrates the transitions between free-energy basins.[28] The free-energy profile is annotated with the following intermolecular hydrogen bond distances (blue): between the oxadiazole and the guanidinium of Arg1173 (top), the NH of the pyrrole and the backbone carbonyl of Pro1110 (middle), and the acetyl oxygen and the amide of the conserved Asn1168 (bottom). The distances were calculated between donor and acceptor atoms, and they were averaged every 4 ns. The two hydrogen bonds involving the 4-acylpyrrole are stable, while the hydrogen bond with the Arg1173 is unstable. (C) The most populated binding mode and two alternative binding modes of compound **1** are indicated with Roman numerals (I, II, and III, respectively).

the pyrrole nitrogen acts as a hydrogen donor to the carbonyl of Pro1110 (in the LPF segment), an interaction first observed in MD simulations of the CREBBP and other bromodomains with the natural acetyllysine ligand.[25] It is interesting to note that this MD simulation result was confirmed by two reports on inhibitors forming a similar hydrogen bond to the conserved proline of various bromodomains.[26]

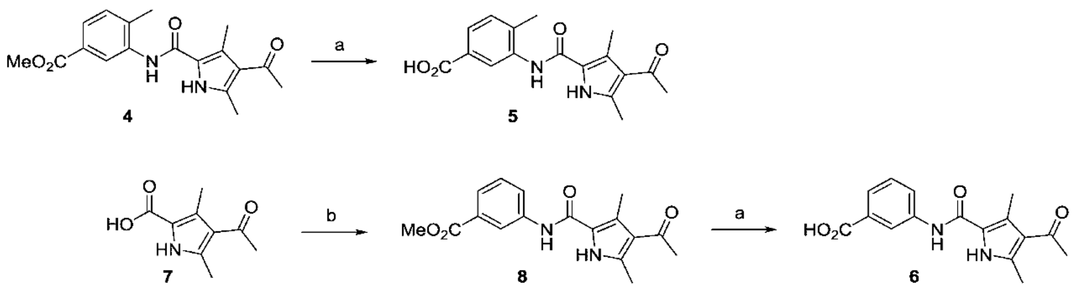
The partially exposed tail group of compound **1** is a 1,3,4-oxadiazole ring, which is in van der Waals contact with the Pro1110 side chain in the LPF shelf. Compound **1** is also involved in favorable polar interactions with the guanidinium group of Arg1173. This positively charged residue is unique to CREBBP (and its paralogue EP300), and thus, interactions with its side chain are likely to result in selectivity toward CREBBP (cf. also refs 4b, 7, 12, and 17). On the contrary, the tail moieties of compounds **2** and **3** (a 4,6-dimethylpyridin-2(1*H*)-one and a 4-ethyl-4*H*-1,2,4-triazole, respectively) are predicted to insert into the so-called ZA channel, which is hydrophobic in the CREBBP bromodomain.

Given the promising K_d value of compound **1** (Figure 6 and Table 1), we aimed to further validate its predicted binding mode into CREBBP by carrying out explicit solvent MD simulations of the complex. Two independent 1 μs MD runs were launched with different initial assignment of the random velocities. The time series of the root-mean-square deviation (RMSD) of compound **1** show that the binding mode is overall stable, although deviations of different magnitude are observed in the two MD runs (Figure 7A). The free-energy profile analysis of the cumulated MD sampling reveals three sub-basins corresponding to three main binding modes, respectively (Figure 7B,C). The largest basin (i.e., pose I) contains about 75% of the 50 000 MD snapshots (saved along 2 μs of total MD sampling) and corresponds to the binding mode predicted by automatic docking. Within this basin, the two hydrogen bonds formed by the 4-acylpyrrole with the conserved Asn and Pro residues are stable. In addition, the 1,3,4-oxadiazole forms a stable interaction with the LPF segment. On the other hand, the direct hydrogen bonds between the 1,3,4-oxadiazole nitrogen atoms and the guanidinium group of Arg1173 are sporadic and show frequent fluctuations during which a single water molecule inserts itself to form a water-mediated hydrogen bond. Quantitatively, the shortest distance between any of the two oxadiazole nitrogen atoms of compound **1** and any NH atoms of the Arg1173 guanidinium is shorter than 3.5 and 5 Å only in 17% and 33% of the MD snapshots, respectively (Figure 7B).

The second largest free-energy basin is sampled in both simulations and is similar to the binding pose predicted by docking of compounds **2** and **3** (compare pose II in Figure 7C to Figure S3). In this alternative pose the 4-acylpyrrole group is anchored in the acetyllysine binding pocket, whereas the 1,3,4-oxadiazole group is in a different orientation with respect to the previous conformation, as it flips toward the ZA channel on the other side of the binding site.

Finally, a third binding conformation was revealed by the free-energy profile (pose III in Figure 7C) with the displace-

Scheme 1[a]



[a]Reagents and conditions: (a) 1 M LiOH, THF, 25 °C, 3–5 h, 59–65%; (b) $(COCl)_2$, DMF, DCM, 30 min at 0 °C and 1 h at 25 °C, then methyl 3-aminobenzoate, Et_3N, DMAP, 25 °C, 12 h, 53%.

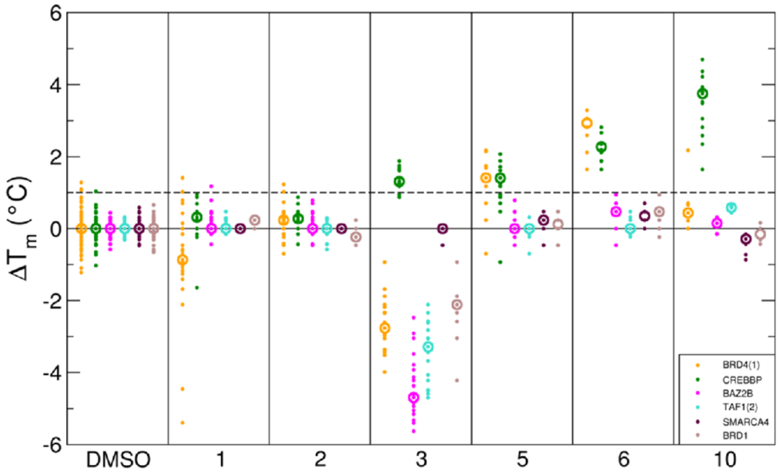
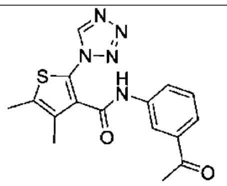
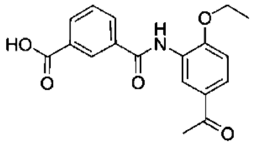


Figure 8. Thermal shift values (ΔT_m) measured in a panel of six bromodomains belonging to six different subfamilies. The plot shows for each compound and each bromodomain the independent measurements (dots) and median value (circle). The dashed line at 1 °C is an arbitrary threshold. The compounds are sorted by compound number along the *x*-axis, and different colors are used for different bromodomains (legend).

Table 2. Experimental Validation of Acylbenzenes Binding to the CREBBP Bromodomain and Selectivity versus BRD4(1) Bromodomain[d]

	Compound	Kd (μM)[a]	LE[b]	Selectivity CREBBP vs. BRD4(1)[c]
9		29	0.24	ND
10		0.77	0.35	>65

[a]K_d values for the CREBBP bromodomain were measured by the competition binding assay[24] (Figure S5). [b]The ligand efficiency has units of kcal/(mol HAC) where HAC is the heavy atom count. [c]Ratio of the K_d values for BRD4(1) and CREBBP bromodomains as measured by the competition binding assay. ND: not determined. [d]Hit **9** was identified by the ALTA procedure, while compound **10** was synthesized.

ment of the tail group away from the LPF segment and the rotation of the acyl group of the pyrrole, which maintains the hydrogen bond with the side chain of the conserved Asn.

The MD analysis of compound **1** suggested that a negatively charged group instead of the 1,3,4-oxadiazole ring might result in more favorable electrostatic interactions with the Arg1173 side chain and thus favor selectivity for CREBBP bromodomain over other bromodomains. Therefore, we decided to carry out a small hit optimization campaign, namely, to replace the 1,3,4-oxadiazole with a carboxylic group. To this end, we synthesized compounds **5**, **6**, and **8** (Scheme 1). Compound **8** was prepared by condensation of commercially available 4-acetyl-3,5-

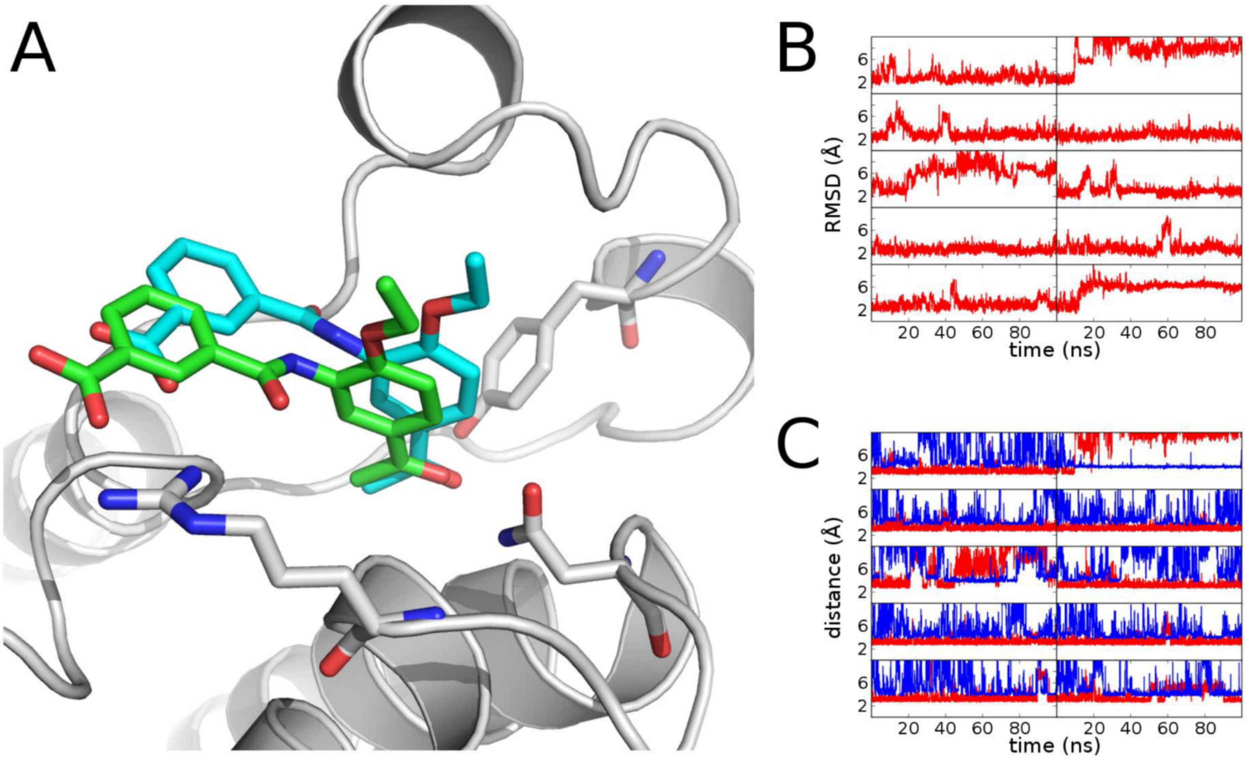


Figure 9. (A) The binding mode of compound **10** predicted by docking and MD simulations (carbon atoms in cyan) is similar to the crystal structure of the complex (carbon atoms in green, from PDB code 4TQN) except for the orientation of the amide linker. Note that both orientations of the amide linker were sampled during the MD simulations but the one observed in the crystal structure was less populated which is likely due to the fact that the MD runs were started with the wrong orientation and the length of the individual runs is too short to reach equilibrium. The crystal structure of CREBBP (gray) is rendered by a ribbon representation with relevant side chains in sticks. (B) Time series of the RMSD of compound **10** along 10 independent MD runs of 100 ns each. The RMSD values were calculated for the heavy atoms of the ligands after overlapping the Cα atoms of the helical segments of the CREBBP bromodomain. (C) Time series of the distance between the carbonyl oxygen of compound **10** and the Asn1168 side chain N atom (red), and the distance between the carboxyl carbon atom of compound **10** and Arg1173 Cζ (blue) along the trajectory.

dimethyl-1*H*-pyrrole-2-carboxylic acid **7** and methyl 3-aminobenzoate. Consequent hydrolysis (in the presence of LiOH) of the methyl ester in **8** and in the commercially available compound **4** afforded the final compounds **5** and **6** in 59% and 65% yield, respectively.

As predicted, the 4-acylpyrrole derivative **6** is more potent than the hit **1** in the competition binding assay and the TR-FRET assay (Table 1). Concerning their selectivity, compounds **5** and **6** show similar thermal shift values for CREBBP and BRD4(1) and no shift for four bromodomains belonging to other subfamilies (Figure 8).

Optimization of Acylbenzene Hit 9 and MD Simulations. Concurrent with the investigations on the 4-acylpyrrole derivatives (compounds **1**–**6**), we optimized the acylbenzene hit **9** as described in ref 17. In short, after purchasing a few commercially available derivatives of **9**, optimization by chemical synthesis guided by analysis of the docking pose resulted in compound **10**, which showed an improved potency (0.77 μM vs 29 μM) and ligand efficiency (0.35 vs 0.24 kcal/mol per non-hydrogen atom) compared to the parent compound **9**. In addition, **10** showed more than 65 times higher affinity for CREBBP than for the first bromodomain of BRD4 (Table 2, Figure S5).

The binding mode of the benzoic acid derivative **10** obtained by docking was first validated by 10 MD simulations of 100 ns each during which the direct salt bridge between the carboxyl group of **10** and the guanidinium of Arg1173 was present for about 70% of the time (up to 80% considering also a single-water bridged interaction, Figure 9B,C). The crystal structure of acylbenzene **10** in complex with the CREBBP bromodomain at 1.7 Å resolution (PDB code 4TQN, Table S1) is in line with the binding mode predicted by docking and MD simulations (Figure 9A).

■ CONCLUSIONS

We have identified two chemical classes of CREBBP bromodomain inhibitors by the ALTA procedure[14] for fragment-based high-throughput docking. The ranking of fragments and molecules was based on the evaluation of the binding energy using a classical force field and including the desolvation penalty calculated in the continuum electrostatics approximation. The hit rate of our in silico screening procedure was high, as two active molecules (K_d values of 13 and 29 μM for ligands **1** and **9**, respectively) were identified by testing in vitro only 17 compounds. Optimization of the two hits was guided by the results of explicit solvent MD simulations.

The 4-acylpyrrole hit **1** underwent a limited optimization because of its lack of selectivity for CREBBP, which is consistent with a recent report on the discovery of 4-acylpyrrole derivatives as BET bromodomain inhibitors.[26a] One of these BET inhibitors (called XD14) was reported to bind also to the CREBBP bromodomain with an affinity of 1.6 μM in the

DOI: 10.1021/acs.jmedchem.5b00171
J. Med. Chem. XXXX, XXX, XXX−XXX

competition binding assay, which is similar to the one of our inhibitor **6**. Of note, our compound **6** has a heavy atom count (HAC) of 22 and thus a more favorable ligand efficiency for CREBBP (0.34 kcal/mol per HAC, Table 1) than the larger XD14 (29 HAC and 0.28 kcal/mol per HAC).

The acylbenzene hit **9** was optimized by chemical synthesis of less than 30 derivatives which are described in detail in ref 17. Briefly, the potency was improved from micromolar to nanomolar with a remarkable increase in ligand efficiency from 0.24 kcal/mol per HAC for the hit **9** to 0.35 kcal/mol per HAC for the derivative **10**. The optimization of both potency and selectivity was based on the favorable polar interactions with the Arg1173 guanidinium of the CREBBP bromodomain as observed in the MD simulations.

In conclusion, both the binding affinity and ligand efficiency of the ALTA hits were improved by the chemical synthesis of a small set of derivatives as in previous applications of the ALTA procedure.[14b,29] MD simulations played a key role, as they provided additional support to the binding mode predicted by docking. Furthermore, the analysis of the interaction motifs along the MD trajectories efficiently guided the hit optimization process.

■ EXPERIMENTAL SECTION

Preparation of the Fragments for Docking. The fragments used in this study were generated from the molecules in the ZINC leads-now library (version of October 2012) by the program DAIM which carries out an automatic decomposition by cleaving all rotatable bonds.[15] The resulting fragments were parametrized according to the CHARMM general force field (CGenFF)[30] using the program MATCH.[31] About 9% of the fragments were lost in this process because the chemical space is not fully covered by this program yet.

Preparation of the Protein. The atomic coordinates of chain A of the 3P1C and 3SVH PDB files were used. Six of the structured water molecules in the receptor binding pocket (water 1, 2, 3, 4, 6, and 7, as defined in ref 32) were kept for docking. PSFGEN was used to add the missing atoms (including hydrogens). Hydrogen atom positions were minimized with CHARMM[20] using the CHARMM27 force field.

Docking of Fragments. The program SEED (solvation energy for exhaustive docking)[16] was used to dock the fragments. SEED calculates the electrostatic energy using a continuum approximation with an efficient numerical evaluation of the Born radii. These radii approximate the space occupied by the solute (i.e., low-dielectric volume) around individual atoms and are employed in the generalized Born formula for the calculation of the electrostatic contribution to the intermolecular interaction and desolvation of the ligand.[33]

Flexible Docking of the Parent Compounds. The parent compounds of the top 3975 (3SVH) and 4601 (3P1C) fragments were retrieved from the ZINC leads-now library. These were docked to the corresponding CREBBP bromodomain structure using AutoDock Vina.[23] All the poses were then refined via minimization by CHARMM.[20] During minimization the interaction energy between the receptor and the compound was evaluated by the sum of the van der Waals and electrostatic energy. The latter energy was computed with a modified Coulombic term, i.e., distance-dependent dielectric constant (4.0r, where r is the distance between partial charges). Then, the poses were filtered by a hydrogen bonding penalty (cutoff 1.0).[21] The remaining compounds were then rescored using the total energy calculated by the program SEED.[16]

The predictive ability of the filtering and scoring of flexible molecules was verified using ROC curves. Seven known inhibitors were used as true positives. ROC curves with AUC values higher than 0.67 were obtained for both structures (Figure S4).

MD Simulations. All MD simulations were carried out by GROMACS 4.6[34] with the CHARMM27/CGenFF force field.[30] The predicted binding poses were used as starting conformations and solvated in cubic water boxes with the TIP3P water model.[35] Long-range electrostatics were treated with the particle mesh Ewald method, while the van der Waals interactions were truncated at a cutoff of 10 Å. After 1 ns of equilibration, production runs were carried out in the *NPT* ensemble at 310 K. The time step was 2 fs, and snapshots were saved for analysis every 20 ps.

Clustering of MD Snapshots and Free-Energy Profile of Binding of Compound 1. A hierarchical tree-based algorithm[36] was used for clustering the coordinate sets saved along the MD simulations. The two 1 μs trajectories were combined, and a total of 50 000 MD snapshots (i.e., one snapshot every 40 ps) were employed for clustering. The RMSD of the ligand heavy atoms and a final fine-grained threshold of 1.0 Å were used for clustering by the tree-based algorithm (with 10 levels of the tree and initial RMSD threshold of 12.0 Å). The tree-based algorithm generated 3929 clusters which were used for calculating the free-energy profile by an automatic procedure which is able to identify all free-energy basins.[27]

Protein Purification and Binding Assays. Protein purification and thermal shift measurements were performed as reported previously.[29b] The description of the competition binding assay and TR-FRET assay is given in the Supporting Information.

■ ASSOCIATED CONTENT

Supporting Information

General procedures for the fragment docking, synthesis and characterization, and biophysical and biological evaluation of final compounds. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.5b00171.

Accession Codes

PDB code for CREBBP in complex with the ligand **10** is 4TQN.

■ AUTHOR INFORMATION

Corresponding Authors

*C.N.: e-mail, cristina.nevado@chem.uzh.ch; phone, +41 44 635 39 45.
*A.C.: e-mail, caflisch@bioc.uzh.ch; phone, +41 44 635 55 21.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Alvaro Salvador for help with chemical synthesis, and Lisa Caflisch, Lisa Gartenmann, and Lars Wiedmer for help with protein purification and thermal shift assay experiments. We thank the Structural Genomics Consortium at Oxford University for providing the plasmids of the bromodomains. The MD simulations were carried out at the Schrödinger computing cluster of the University of Zurich. This work was supported financially by the Swiss National Science Foundation and the Swiss Cancer League (Krebsliga). D.S. is a recipient of the SystemsX.ch translational postdoc fellowship.

■ ABBREVIATIONS USED

ALTA, anchor-based library tailoring; AUC, area under the curve; BET, bromodomain and extraterminal; CHARMM, chemistry at Harvard molecular mechanics; CREBBP, CREB binding protein; DMSO, dimethyl sulfoxide; HAC, heavy atom count; MD, molecular dynamics; NMR, nuclear magnetic resonance; RMSD, root-mean-square deviation; ROC, receiver operating characteristic; SEED, solvation energy for exhaustive docking; SEM, standard error of the mean; TR-FRET, time-resolved Förster resonance energy transfer

■ REFERENCES

(1) Sanchez, R.; Meslamani, J.; Zhou, M. M. The bromodomain: from epigenome reader to druggable target. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2014**, *1839*, 676−685.

(2) Filippakopoulos, P.; Knapp, S. The bromodomain interaction module. *FEBS Lett.* **2012**, *586*, 2692−2704.

(3) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012**, *149*, 214−231.

(4) (a) Sachchidanand; Resnick-Silverman, L.; Yan, S.; Mutjaba, S.; Liu, W.; Zeng, L.; Manfredi, J. J.; Zhou, M.-M. Target Structure-Based Discovery of Small Molecules that Block Human p53 and CREB Binding Protein Association. *Chem. Biol.* **2006**, *13*, 81−90. (b) Borah, J. C.; Mujtaba, S.; Karakikes, I.; Zeng, L.; Muller, M.; Patel, J.; Moshkina, N.; Morohashi, K.; Zhang, W.; Gerona-Navarro, G.; Hajjar, R. J.; Zhou, M. M. A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. *Chem. Biol.* **2011**, *18*, 531−541.

(5) Gerona-Navarro, G.; Yoel, R.; Mujtaba, S.; Frasca, A.; Patel, J.; Zeng, L.; Plotnikov, A. N.; Osman, R.; Zhou, M. M. Rational design of cyclic peptide modulators of the transcriptional coactivator CBP: a new class of p53 inhibitors. *J. Am. Chem. Soc.* **2011**, *133*, 2040−2043.

(6) Philpott, M.; Yang, J.; Tumber, T.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Keates, T.; Felletar, I.; Ciulli, A.; Knapp, S.; Heightman, T. D. Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. BioSyst.* **2011**, *7*, 2899−2908.

(7) Hewings, D. S.; Wang, M.; Philpott, M.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Vuppusetty, C.; Marsden, B.; Knapp, S.; Conway, S. J.; Heightman, T. D. 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J. Med. Chem.* **2011**, *54*, 6761−6770.

(8) Chung, C.-w.; Dean, A. W.; Woolven, J. M.; Bamborough, P. Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* **2012**, *55*, 576−586.

(9) Hay, D.; Fedorov, O.; Filippakopoulos, P.; Martin, S.; Philpott, M.; Picaud, S.; Hewings, D. S.; Uttakar, S.; Heightman, T. D.; Conway, S. J.; Knapp, S.; Brennan, P. E. The design and synthesis of 5- and 6-isoxazolylbenzimidazoles as selective inhibitors of the BET bromodomains. *MedChemComm* **2013**, *4*, 140−144.

(10) Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Muller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc.* **2014**, *136*, 9308−9319.

(11) I-CBP112—a CREBBP/EP300-selective chemical probe. http://www.thesgc.org/chemical-probes/ICBP112.

(12) Rooney, T. P.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin, S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Muller, S.; Knapp, S.; Brennan, P. E.; Conway, S. J. A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation-pi interaction. *Angew. Chem., Int. Ed.* **2014**, *53*, 6126−6130.

(13) Fedorov, O.; Lingard, H.; Wells, C.; Monteiro, O. P.; Picaud, S.; Keates, T.; Yapp, C.; Philpott, M.; Martin, S. J.; Felletar, I.; Marsden, B. D.; Filippakopoulos, P.; Muller, S.; Knapp, S.; Brennan, P. E. [1,2,4]triazolo[4,3-a]phthalazines: inhibitors of diverse bromodomains. *J. Med. Chem.* **2014**, *57*, 462−476.

(14) (a) Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors. *Proteins: Struct., Funct., Genet.* **2008**, *73*, 11−18. (b) Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a Novel Chemotype of Tyrosine Kinase Inhibitors by Fragment-Based Docking and Molecular Dynamics. *ACS Med. Chem. Lett.* **2012**, *3*, 834−838.

(15) Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking. *J. Med. Chem.* **2006**, *49*, 7384−7392.

(16) (a) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive Docking of Molecular Fragments With Electrostatic Solvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88−105. (b) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 256−268.

(17) Unzue, A.; Xu, M.; Dong, J.; Wiedmer, L.; Spiliotopoulos, D.; Caflisch, A.; Nevado, C. Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b00172.

(18) Steiner, S.; Magno, A.; Huang, D.; Caflisch, A. Does bromodomain flexibility influence histone recognition? *FEBS Lett.* **2013**, *587*, 2158−2163.

(19) Chung, C.-w.; Coste, H.; White, J. H.; Mirguet, O.; Wilde, J.; Gosmini, R. L.; Delves, C.; Magny, S. M.; Woodward, R.; Hughes, S. A.; Boursier, E. V.; Flynn, H.; Bouillot, A. M.; Bamborough, P.; Brusq, J.-M. G.; Gellibert, F. J.; Jones, E. J.; Riou, A. M.; Homes, P.; Martin, S. L.; Uings, I. J.; Toum, J.; Clément, C. A.; Boullay, A.-B.; Grimley, R. L.; Blandel, F. M.; Prinjha, R. K.; Lee, K.; Kirilovsky, J.; Nicodeme, E. Discovery and Characterization of Small Molecule Inhibitors of the BET Family Bromodomains. *J. Med. Chem.* **2011**, *54*, 3827−3838.

(20) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(21) Zhao, H.; Huang, D. Hydrogen Bonding Penalty upon Ligand Binding. *PLoS One* **2011**, *6*, e19923.

(22) Filippakopoulos, P.; Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discovery* **2014**, *13*, 337−356.

(23) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(24) (a) Fabian, M. A.; Biggs, W. H., 3rd; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329−336. (b) Quinn, E.; Wodicka, L.; Ciceri, P.; Pallares, G.; Pickle, E.; Torrey, A.; Hunt, J.; Treiber, D. BROMOscan - a high throughput, quantitative ligand binding platform identifies best-in-class bromodomain inhibitors from a screen of mature compounds targeting other protein classes. *Cancer Res.* **2013**, *73*, 4238.

(25) Magno, A.; Steiner, S.; Caflisch, A. Mechanism and Kinetics of Acetyl-Lysine Binding to Bromodomains. *J. Chem. Theory Comput.* **2013**, *9*, 4225−4232.

(26) (a) Lucas, X.; Wohlwend, D.; Hügle, M.; Schmidtkunz, K.; Gerhardt, S.; Schüle, R.; Jung, M.; Einsle, O.; Günther, S. 4-Acyl Pyrroles: Mimicking Acetylated Lysines in Histone Code Reading. *Angew. Chem., Int. Ed.* **2013**, *52*, 14055−14059. (b) Ferguson, F. M.; Fedorov, O.; Chaikuad, A.; Philpott, M.; Muniz, J. R. C.; Felletar, I.; von Delft, F.; Heightman, T.; Knapp, S.; Abell, C.; Ciulli, A. Targeting low druggability bromodomains: Fragment based screening and inhibitor design against the BAZ2B bromodomain. *J. Med. Chem.* **2013**, *56*, 10183−101837.

(27) Blöchliger, N.; Vitalis, A.; Caflisch, A. A scalable algorithm to order and annotate continuous observations reveals the metastable

states visited by dynamical systems. *Comput. Phys. Commun.* **2013**, *184*, 2446−2453.

(28) Blöchliger, N.; Vitalis, A.; Caflisch, A. High-resolution visualisation of the States and pathways sampled in molecular dynamics simulations. *Sci. Rep.* **2014**, *4*, 6264.

(29) (a) Lafleur, K.; Huang, D.; Zhou, T.; Caflisch, A.; Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **2009**, *52*, 6433−6446. (b) Zhao, H.; Gartenmann, L.; Dong, J.; Spiliotopoulos, D.; Caflisch, A. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2493−2496.

(30) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(31) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **2012**, *33*, 189−202.

(32) Huang, D.; Rossini, E.; Steiner, S.; Caflisch, A. Structured Water Molecules in the Binding Site of Bromodomains Can Be Displaced by Cosolvent. *ChemMedChem* **2014**, *9*, 573−579.

(33) (a) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098−8106. (b) Scarsi, M.; Apostolakis, J.; Caflisch, A. Comparison of a GB Solvation Model with Explicit Solvent Simulations: Potentials of Mean Force and Conformational Preferences of Alanine Dipeptide and 1,2-Dichloroethane. *J. Phys. Chem. B* **1998**, *102*, 3637−3641.

(34) (a) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447. (b) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(35) Jorgensen, W. L.; Madura, J. D. Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. *J. Am. Chem. Soc.* **1983**, *105*, 1407−1413.

(36) Vitalis, A.; Caflisch, A. Efficient Construction of Mesostate Networks from Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2012**, *8*, 1108−1120.

Discovery of CREBBP Bromodomain Inhibitors by High-throughput Docking and Hit Optimization Guided by Molecular Dynamics

Min Xu,[a] Andrea Unzue,[b] Jing Dong, [a] Dimitrios Spiliotopoulos, [a] Cristina Nevado*[b], and Amedeo Caflisch*[a]

[a]M. Xu, Dr. J. Dong, Dr. D. Spiliotopoulos, Prof. Dr. A. Caflisch* Department of Biochemistry, Universität Zürich. Winterthurerstrasse 190, CH-8057. Zürich, Switzerland.

[b]A. Unzue, Prof. Dr. C. Nevado* Department of Chemistry, Universität Zürich. Winterthurerstrasse 190, CH-8057. Zürich, Switzerland.

Supplementary data

Table of contents

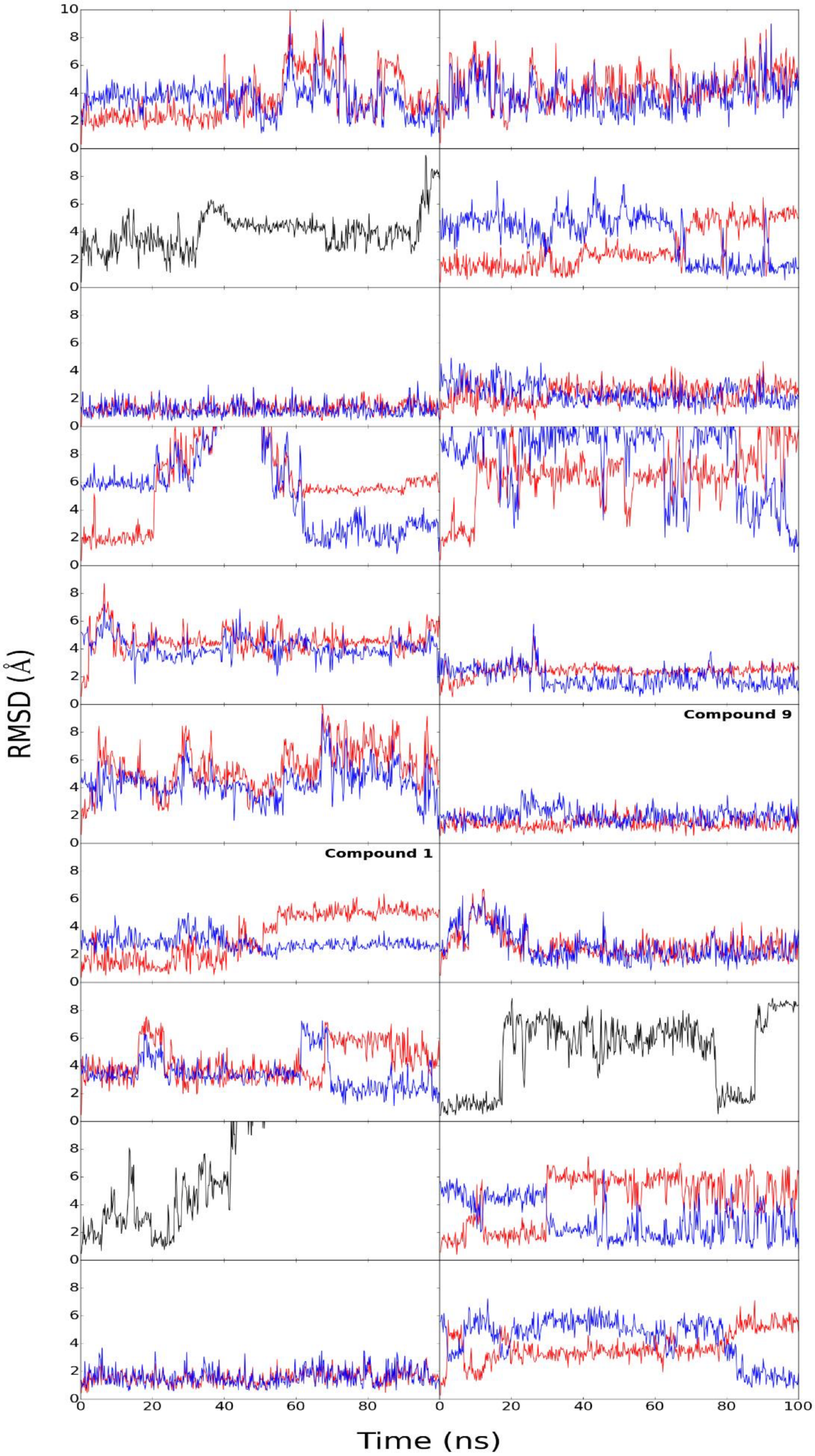


Figure S1. Time series of the RMSD of the heavy atoms of the compound from the starting structure, i.e., the docked pose (red) and from the final snapshot (blue) along the 100-ns MD runs of the top 20 compounds suggested by docking. The RMSD time-series were calculated after overlapping the rigid part of the protein. Full unbinding was observed for three compounds (black curves); these compounds were not purchased.

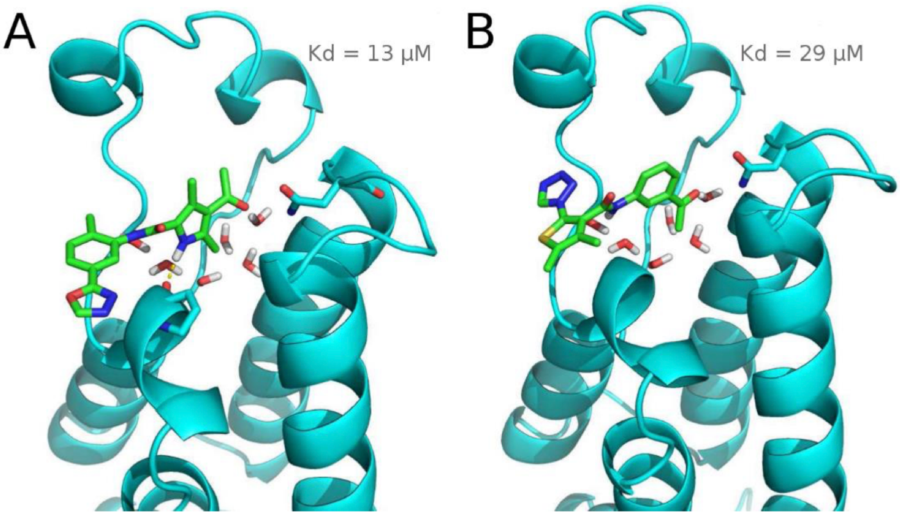


Figure S2. Docked pose of compound **1** in the 3P1C structure of CREBBP (panel A) and compound **9** in the 3SVH structure of CREBBP (panel B). The ligands are shown in sticks (carbon atoms in green) while the CREBBP bromodomain is represented by a ribbon model (cyan) with sticks showing the side chain of the conserved Asn1168 and the conserved water molecules. The K_d values were obtained by a competition binding assay (at DiscoveRx).

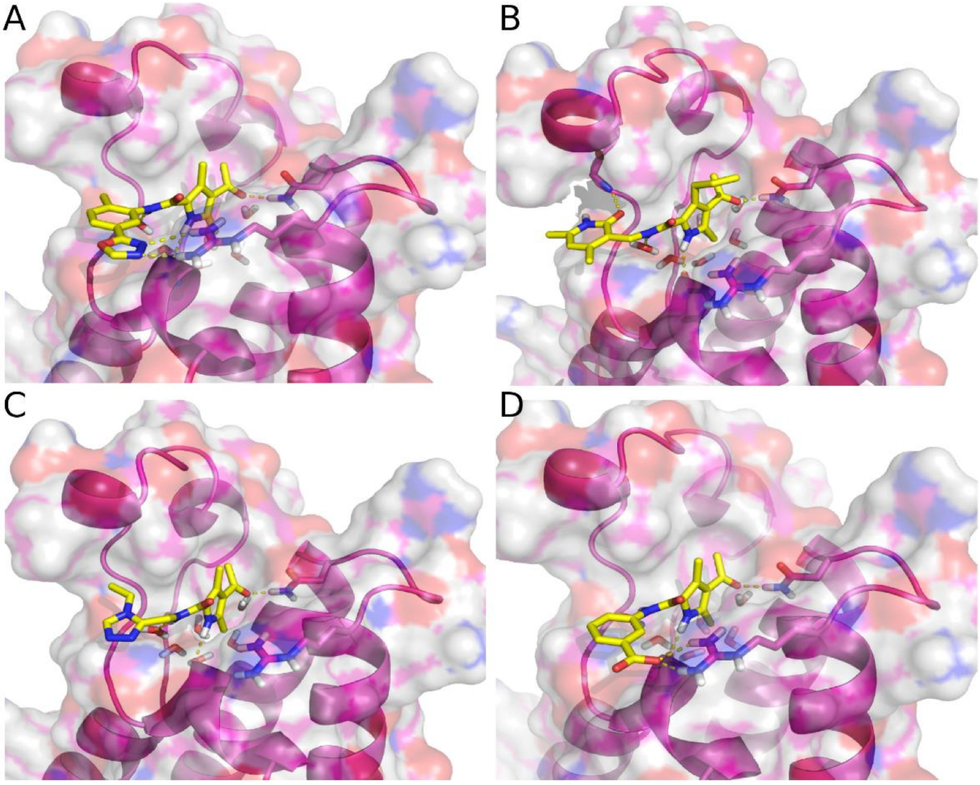


Figure S3. Binding pose of compounds **1** (panel A), **2** (B), **3** (C), and **6** (D) as predicted by docking. The protein backbone is shown by a ribbon model and transparent surface with sticks for the side chains of the conserved Asn1168 and Arg1173. Hydrogen bonds are shown by dashed lines.

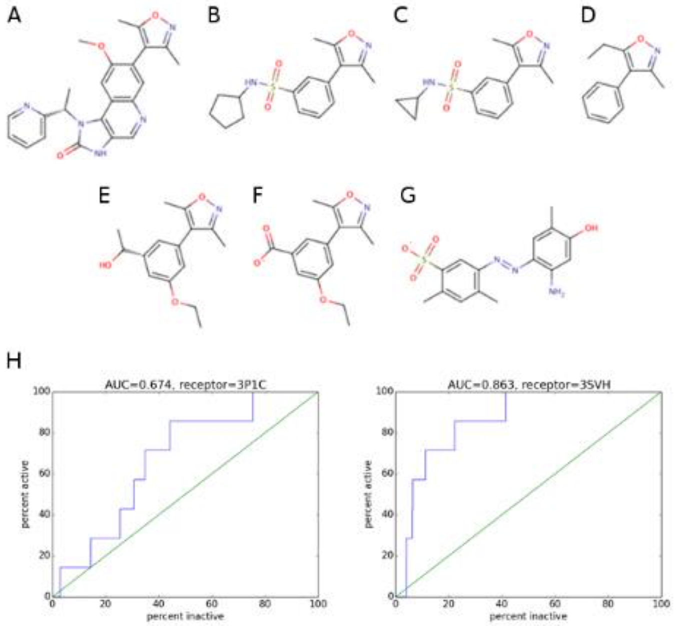


Figure S4. Seven known inhibitors of CREBBP (A-G) used as true positives for the ROC plots (H). (A) From PDB code 3ZYU,[1] (B) from PDB code 4A9N, (C) from PDB code 4A9M, (D) from PDB code 4A9O,[2] (E, F), [3] and (G) From PDB code 2L84.[4] (H) The two ROC plots were calculated using the docking results on the PDB structures 3P1C (left) and 3SVH (right).

Experimental section

TR-FRET assays

TR-FRET assays were carried out in duplicate at BPS Bioscience using a recombinant CBP bromodomain (BPS catalogue #31128) and the BET Ligand (BPS catalogue #33000) as provided in the CREBBP TR-FRET Assay Kit (BPS catalogue #32619). A 10 mM solution of the compound under investigation in DMSO was prepared and shipped to BPS Bioscience, where it was tested at 10 concentrations over the range of 0.001-10 µM (compounds **5** and **6**) or 0.01-100 µM (compound **1**) . Each compound solution was then diluted in water to obtain a 10% DMSO solution. 2 µL of this dilution were added to a 20 µL reaction mixture (12.5 nM CBP, 125 nM BET Ligand, including FRET dyes and the amount of compound needed to reach the required concentration). The resulting mixture was incubated for 2 hours at room temperature prior to reading the TR (time resolved)-FRET signal using a Tecan Infinite M1000 plate reader. The negative control consisted of the aforementioned mixture in which the buffer was added in place of compound. TR-FRET were recorded as the ratio of the fluorescence of the acceptor and the donor dyes (acceptor/donor).

The TR-FRET data was analyzed using Graphpad Prism software. The percent activity in the presence of each compound was calculated according to the following equation: % activity = [(F- Fb)/(Ft – Fb)]×100, where Ft is the TR-FRET signal in the absence of any compound (100 % activity), Fb the TR-FRET signal in the absence of the bromodomain (0 % activity) and F the TR-FRET signal in the presence of the compound. The percent inhibition was calculated according to the following equation: % inhibition = 100 - % activity. The values of % activity versus a series of compound concentrations were then plotted using non-linear regression analysis of Sigmoidal dose-response curve generated with the equation $Y=B+(T-B)/1+10^{((LogIC50-X)\times Hill\ Slope)}$, where Y=percent activity, B=minimum percent activity, T=maximum percent activity, X= logarithm of compound and Hill Slope=slope factor or Hill coefficient. The IC_{50} value corresponds to the concentration causing a half-maximal percent activity.

BROMOscan assays

K_D determinations by means of BROMOscan technology was carried out at DiscoveRx. E. *coli* derived from BL21 strain was used as host to grow T7 phage strains displaying the bromodomains. E. *coli*, grown to log-phase, were infected with T7 phage (from a frozen stock, being the multiplicity of infection 0.4) and incubated while shaking at 32 °C for 90-150 minutes, until lysis. In order to remove cell debris, lysates were centrifuged at 5,000 x g and filtered (0.2 µm). Affinity resins were obtained by treating streptavidin-coated magnetic beads with biotinylated acetylated peptide ligands for 30 minutes at 25°C. Those beads were then blocked with excess of biotin and washed with blocking buffer (SeaBlock (Pierce), 1% bovine serum albumin, BSA, 0.05% Tween20, 1 mM dithiothreitol, (DTT) removing the unbound ligand and reducing non-specific phage binding.

During the experiment, the bromodomain, ligand-bound affinity beads and test compounds were combined in a buffer composed of 17% SeaBlock, 0.33x phosphate-buffered solution, PBS, 0.04% Tween20, 0.02% BSA, 0.004% sodium azide and 7.4 mM DTT. Test compounds were prepared as 50 mM in pure DMSO and diluted to 5 mM with monoethylene gycol, MEG (100× concentrated in respect to the top screening concentration, 50 µM). During the assay a DMSO and MEG final concentration of 0.1% and 0.9% respectively was used. The assays were carried out in polystyrene 96-well plates in a final volume of 0.135 mL. The assay plates were incubated at 25 °C with shaking for 1 hour and the affinity beads were washed with a buffer composed of 0.05% Tween 20 in PBS. The beads were then re-suspended in the elution buffer (1x PBS, 0.05% Tween 20, 2 µM non-biotinylated affinity ligand) and incubated at 25°C with shaking for 30 minutes. The bromodomain concentration in the eluates was measured by qPCR. Binding constants (K_d) were calculated with a standard dose-response curve using the Hill equation and curves were fitted using a non-linear least square fit with the Levenberg-Marquardt algorithm.

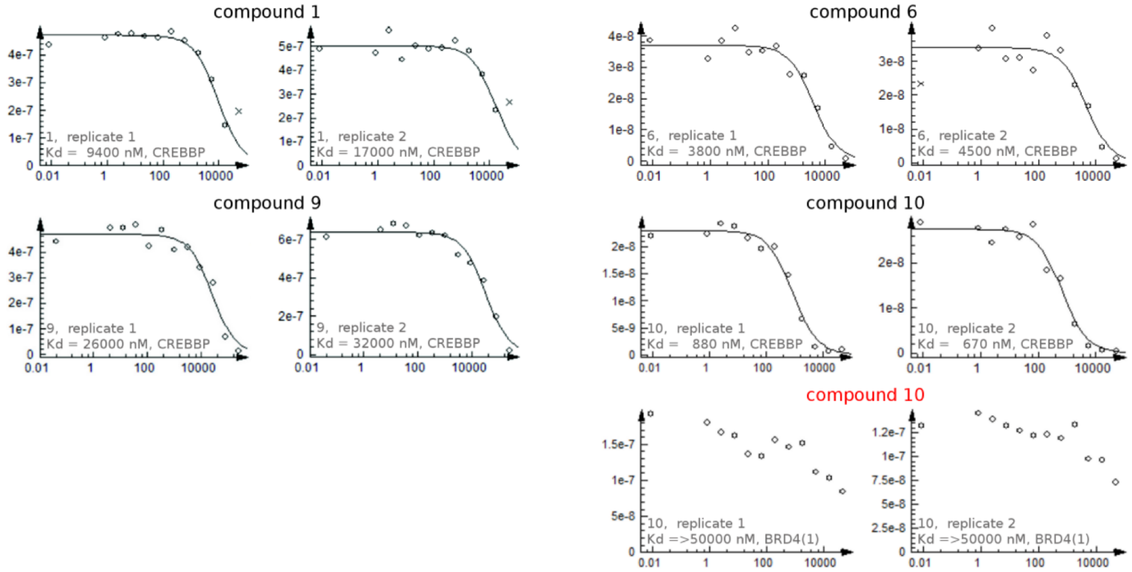


Figure S5. Dose-response curves for compounds **1**, **6**, **9,** and **10** obtained using the competition binding assay at DiscoveRx with CREBBP (black), and BRD4(1) only for compound **10** (red).

Crystallization, Data Collection, and Structure Determination

Crystals of the CREBBP bromodomain were grown at 4°C using the hanging drop vapor diffusion method. A 50 mM solution of compound **10** (in 100% DMSO) was added to a solution of CREBBP to reach a final DMSO concentration of 1% (v/v) and the mixture was incubated on ice for 1 hour before crystallization. Then equal volumes of protein (with compound **10**) and reservoir solutions (0.1 M MES pH 6.5, 0.10 MgCl2, 20 % PEG 6000, 10 % ethylene glycol) were mixed and crystals appeared after 1 to 2 days. The crystals were flash-frozen in liquid nitrogen with extra 10% ethylene glycol as cryoprotectant for measurements. Data sets were collected on a PILATUS 6MF detector at the Swiss Light Source beamline X06SA of the Paul Scherrer Institute (Villigen, Switzerland) and indexed, integrated and scaled with the XDS[5] and CCP4 programs.[6] The structures were solved by molecular replacement with PHASER[7] using the CREBBP structure (PDB entry 4NR5) as a search model and refined with PHENIX.[8] The atomic coordinates and structure factors of CREBBP in complex with inhibitor **10** have been deposited with the Protein Data Bank as entry 4TQN. Details are shown in Table S1.

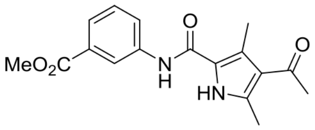
Table S1

	PDB code 4TQN
Space group	P1 21 1
Unit cell	
a (Å)	24.94
b (Å)	42.94
c (Å)	51.98
alpha	90.00
beta	97.24
gamma	90.00
Resolution range (Å)	42.94 -1.70
Unique reflections	12119(1768)
<I/σ(I)>	15.6(5.3)
R merge	0.068(0.380)
Completeness (%)	99.9(99.2)
Multiplicity	6.5(6.2)
Refinement	
Resolution range (Å)	33.00-1.70
R factor/R free	0.1813/0.1990
Mean B factors (A2)	23.24
RMS bonds (Å)	0.006
RMS angles (°)	1.155

Chemistry

All reactions, unless otherwise stated, were carried out under a nitrogen atmosphere using standard Schlenk-techniques. All reagents were used as received unless otherwise noted. Solvents were purchased in the best quality available, degassed by purging thoroughly with nitrogen and dried over activated molecular sieves of appropriate size. Alternatively, they were purged with argon and passed through alumina columns in a solvent purification system (Innovative Technology). Reactions were monitored by thin layer chromatography (TLC) using Merck TLC silica gel 60 F_{254}. Flash column chromatography was performed over silica gel (230-400 mesh). NMR spectra were recorded on AV2 500 MHz Bruker spectrometers. Chemical shifts are given in ppm. The spectra are calibrated to the residual ^1H and ^{13}C signals of the solvents. Multiplicities are abbreviated as follows: singlet (s), doublet (d), triplet (t), quartet (q), doublet-doublet (dd), quintet (quint), septet (sept), multiplet (m), and broad (br). High-resolution electrospray ionization mass spectrometry was performed on a Finnigan MAT 900 (Thermo Finnigan, San Jose, CA, USA) double-focusing magnetic sector mass spectrometer. Ten spectra were acquired. A mass accuracy ≤2 ppm was obtained in the peak matching acquisition mode by using a solution containing 2 µL PEG200, 2 µL PPG450, and 1.5 mg NaOAc (all obtained from Sigma-Aldrich, Buchs, Switzerland) dissolved in 100 mL MeOH (HPLC Supra grade, Scharlau, E-Barcelona) as internal standard. The purity of all tested compounds was determined by HPLC on a Waters Acquity UPLC (Waters, Milford, MA) Top spectrometer using an Acquity BEH C18 HPLC column (1.7 µm, 1× 50 mm, Waters) with a mixture of H_2O + 0.1% HCOOH (A) and CH_3CN + 0.1% HCOOH (B) solvent (0.1 mL flow rate, linear gradient from 5% to 98% B within 4 min followed by flushing with 98% B for 1 min). Unless otherwise stated, all compounds showed ≥ 95 % purity.

Methyl 3-(4-acetyl-3,5-dimethyl-1*H*-pyrrole-2-carboxamido)benzoate (8)

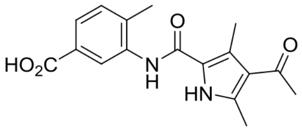


To a solution of 4-acetyl-3,5-dimethyl-1H-pyrrole-2-carboxylic acid (50 mg, 0.28 mmol) in DCM (0.5 mL) at 0 °C oxalyl chloride (24 µL, 0.28 mmol) and one drop of DMF were added. The reaction mixture was stirred at 0 °C for 30 min followed by stirring at 25 °C for 1 h. Methyl 3-aminobenzoate (35 mg, 0.23 mmol), Et_3N (77 µL, 0.55 mmol) and DMAP (4.0 mg, 0.033 mmol) were then added at 0 °C and the reaction was allowed to warm to 25 °C and stirred for 12 h. The reaction mixture was quenched with saturated NH_4Cl solution and extracted three times with EtOAc. The combined organic phases were dried over $MgSO_4$ and concentrated under reduced pressure. The residue was then purified by flash column chromatography (hexane/EtOAc, 1:1) affording the desired amide in pure form (38 mg, 53 %). Off white solid; mp 190-195 °C; ^1H NMR (500 MHz, DMSO-d_6): δ = 11.64 (s, 1H), 9.76 (s, 1H), 8.30 (s, 1H), 7.96 (d, *J* = 8.1 Hz, 1H), 7.66 (d, *J* = 7.7 Hz, 1H), 7.48 (t, *J* = 7.9 Hz, 1H), 3.87 (s, 3H), 2.49 (s, 3H), 2.49 (s, 3H), 2.38 (s, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ = 194.3, 166.1, 159.8, 139.6, 136.9, 130.0, 129.2, 125.8, 124.1, 123.7, 122.2, 121.4, 120.1, 52.2, 31.1, 14.6, 12.4; IR (neat): ṽ = 3248, 2958, 2925, 1719, 1626, 1591, 1553, 1534, 1506, 1476, 1439, 1428, 1414, 1300, 1284, 1261, 1236, 1177, 1095, 1080, 1039, 948, 804, 782, 773, 749 cm^{-1}; HRMS (ESI), *m/z*: calcd for $C_{17}H_{18}N_2NaO_4^+$, 337.1159; found, 337.1158.

General procedure for ester hydrolysis

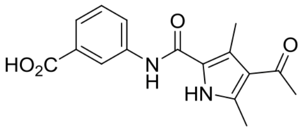
To a solution of the methyl ester (1 eq) in THF (0.1 M) 1M LiOH solution (5 eq) was added. The reaction mixture was stirred at 25 °C for 3-5 h. The reaction mixture was concentrated under reduced pressure and 1M HCl was added. The obtained precipitate was washed with hexanes, ether and cold DCM, affording the desired carboxylic acids in pure form.

3-(4-acetyl-3,5-dimethyl-1*H*-pyrrole-2-carboxamido)-4-methylbenzoic acid (5)



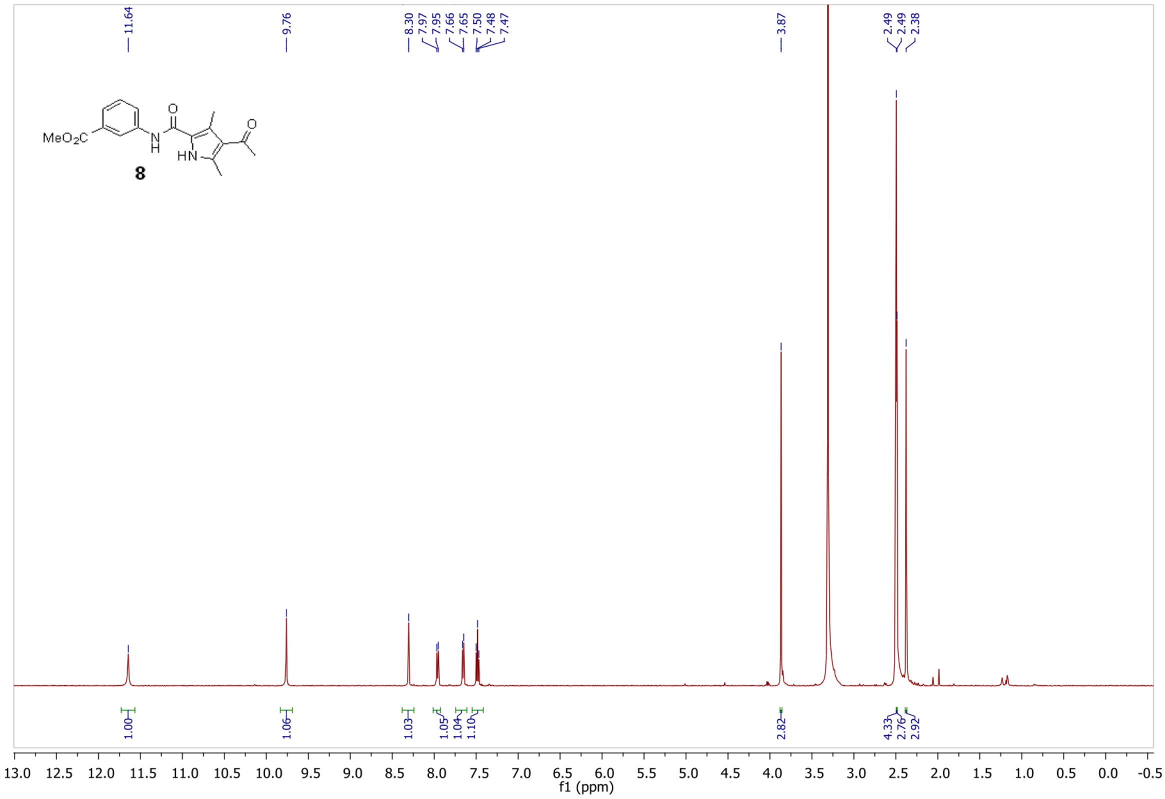
Off white solid; Yield: 59%; mp 267-270 °C; ^{1}H NMR (500 MHz, DMSO-d_6): δ = 12.88 (br, 1H), 11.74 (s, 1H), 9.17 (s, 1H), 8.05 (s, 1H), 7.69 (d, *J* = 7.7 Hz, 1H), 7.38 (d, *J* = 8.0 Hz, 1H), 2.52 (s, 3H), 2.48 (s, 3H), 2.48 (s, 3H), 2.30 (s, 3H); ^{13}C NMR (126 MHz, DMSO-d_6): δ = 194.5, 167.1, 160.0, 137.9, 137.0, 136.7, 130.7, 128.8, 126.1, 124.8, 122.1, 121.8, 31.2, 18.4, 14.6, 12.6, one C is missing due to overlapping; IR (neat): ῦ = 3451, 3257, 3018, 1719, 1617, 1580, 1540, 1480, 1457, 1412, 1367, 1312, 1280, 1254, 1224, 1189, 1175, 1075, 1040, 989, 919, 805, 765, 743 cm^{-1}; HRMS (ESI), *m*/*z*: calcd for $C_{17}H_{18}N_2NaO_4^+$, 337.1159; found, 337.1159.

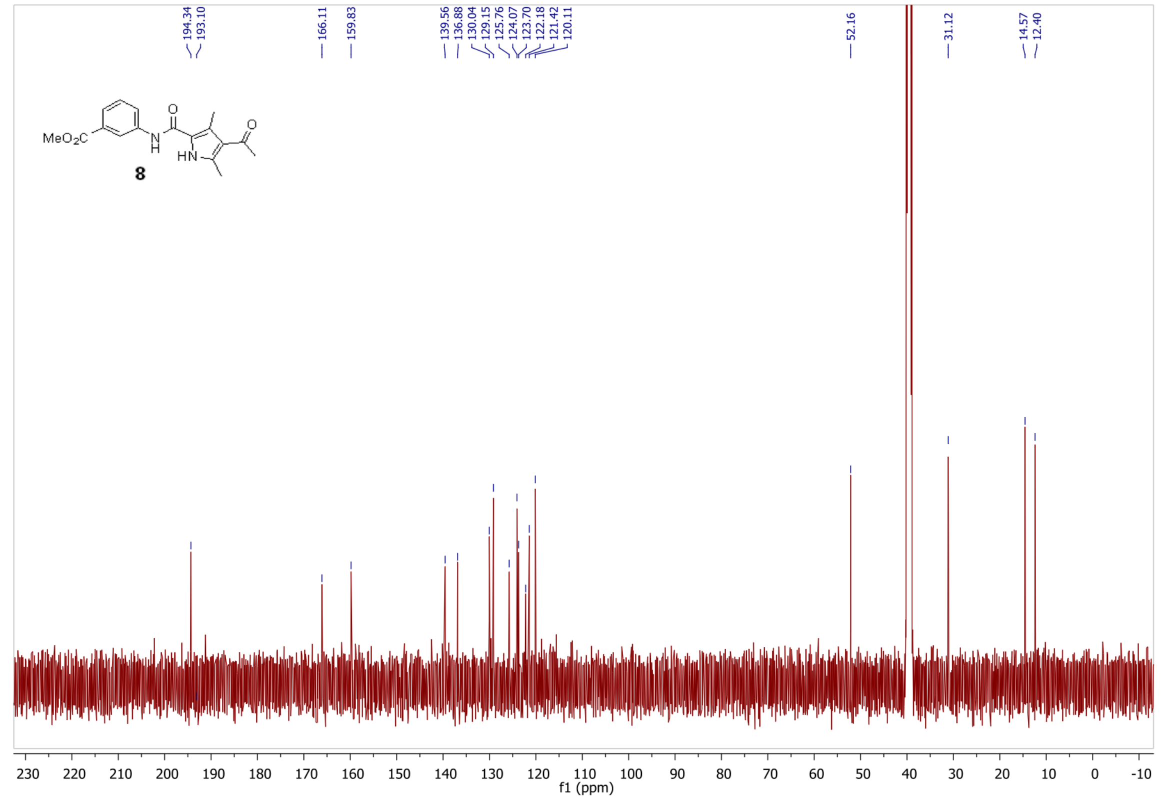
3-(4-acetyl-3,5-dimethyl-1*H*-pyrrole-2-carboxamido)benzoic acid (6)

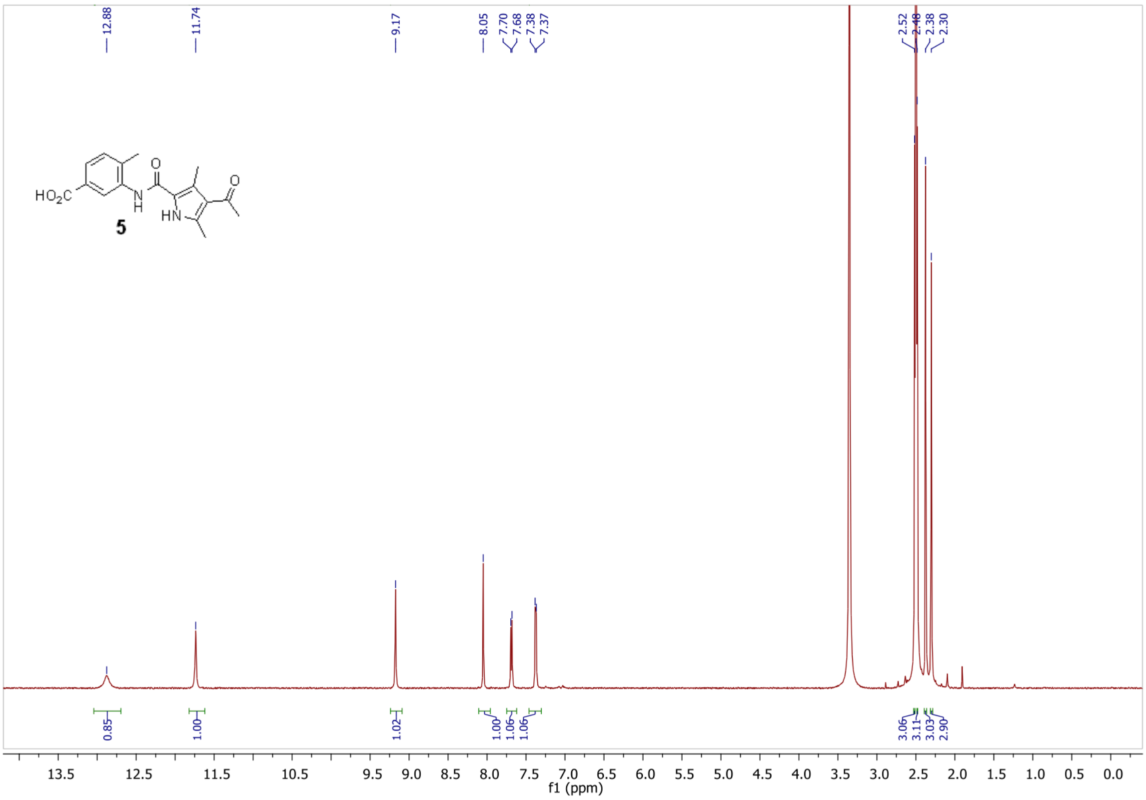


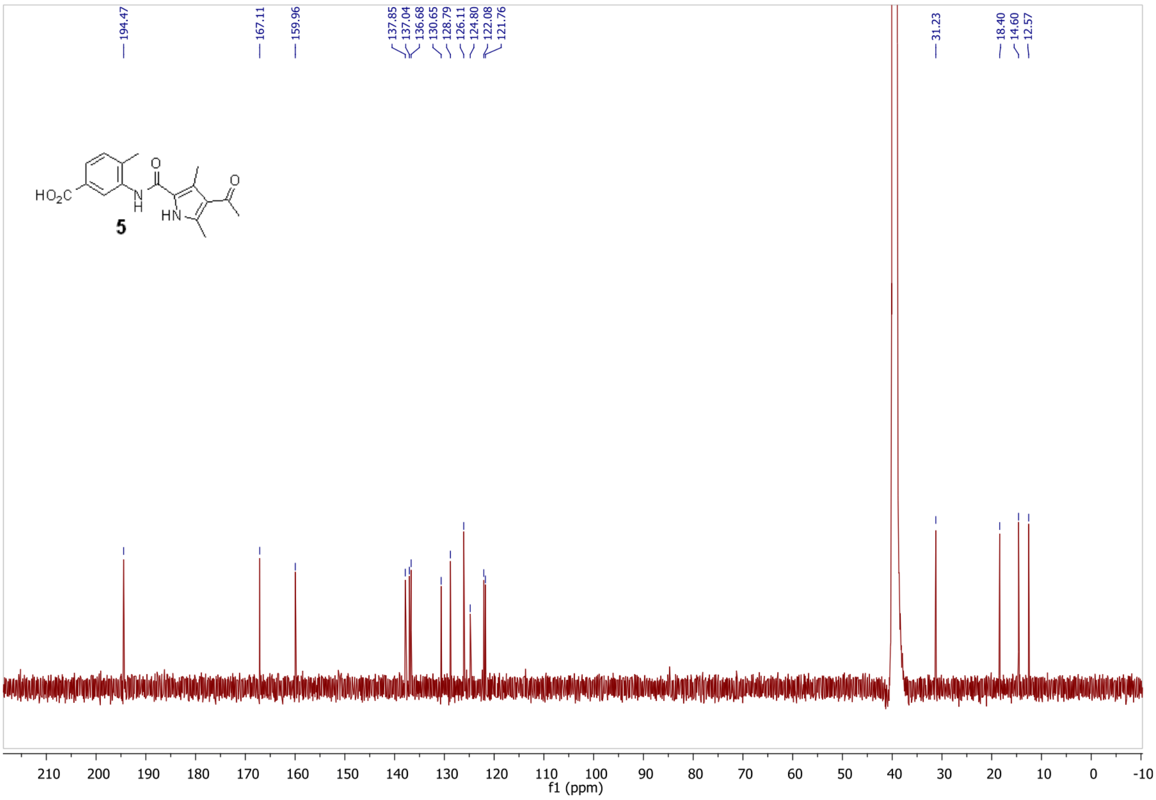
Pale brown solid; Yield: 65%; mp 249-253 °C; ^{1}H NMR (500 MHz, DMSO-d_6): δ = 10.03 (br, 1H), 8.28 (s, 1H), 7.99 (s, 1H), 7.62 (d, *J* = 7.5 Hz, 1H), 7.40 (t, *J* = 7.4 Hz, 1H), 2.37 (s, 3H), six H are missing due to overlapping with the solvent; ^{13}C NMR (126 MHz, DMSO-d_6): δ = 194.9, 160.3, 139.9, 137.6, 129.1, 126.6, 124.3, 122.6, 121.9, 121.0, 31.7, 15.2, 13.0, three C are missing due to overlapping; IR (neat): ῦ = 3441, 3182, 3109, 1704, 1626, 1591, 1542, 1483, 1439, 1413, 1363, 1294, 1250, 1186, 1069, 1038, 842, 806, 751 cm^{-1}; HRMS (ESI), *m*/*z*: calcd for $C_{16}H_{16}N_2NaO_4^+$, 323.1002; found, 323.0999.

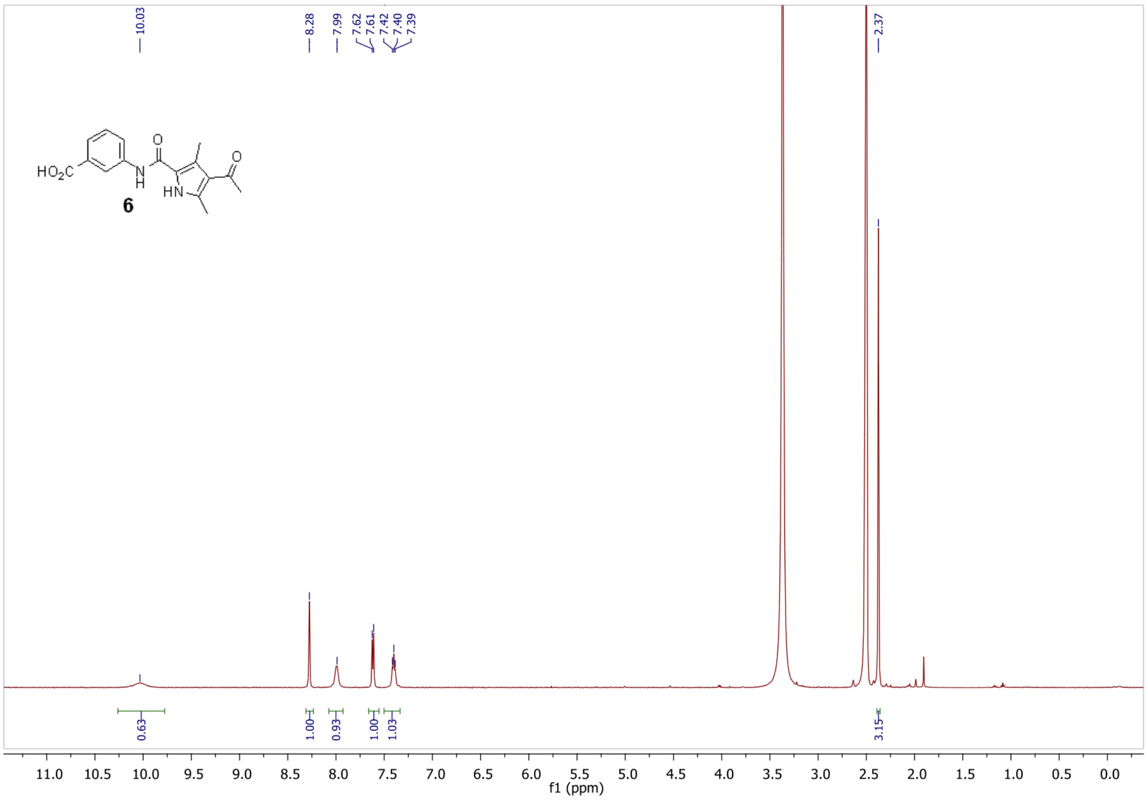
^{1}H-NMR and ^{13}C-NMR spectra

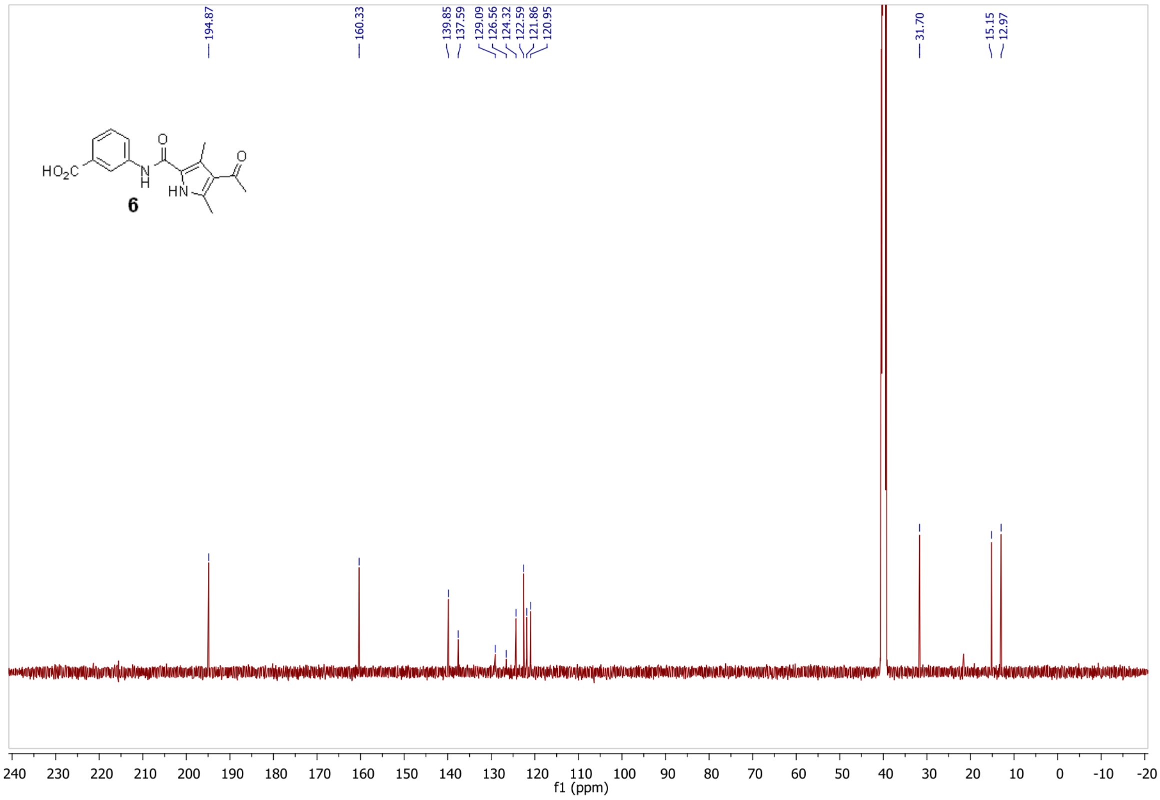




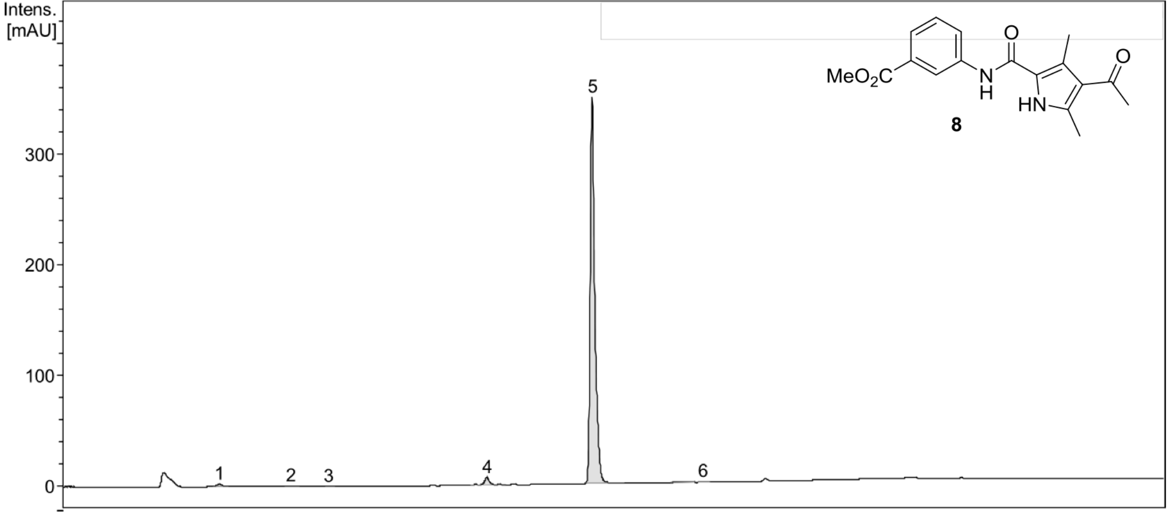

12.88
11.74
9.17
8.05
7.70
7.68
7.38
7.37
2.52
2.48
2.38
2.30
HO2C
O
N
H
HN
O
5
0.85
1.00
1.02
1.00
1.06
1.06
3.06
3.11
3.03
2.90
13.5
12.5
11.5
10.5
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)


194.47
167.11
159.96
137.85
137.04
136.68
130.65
128.79
126.11
124.80
122.08
121.76
31.23
18.40
14.60
12.57
HO2C
O
N
H
HN
O
5
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

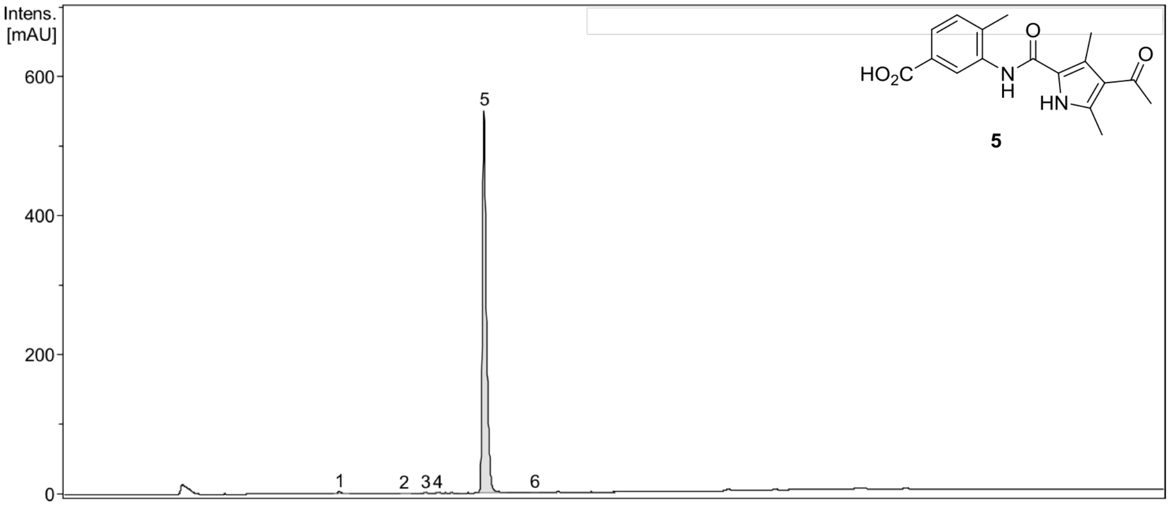

10.03
8.28
7.99
7.62
7.61
7.42
7.40
7.39
2.37
HO2C
N
H
O
HN
O
6
0.63
1.00
0.93
1.00
1.03
3.15
11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)


194.87
160.33
139.85
137.59
129.09
126.56
124.32
122.59
121.86
120.95
31.70
15.15
12.97
HO2C
N
H
O
HN
O
6
240 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20
f1 (ppm)

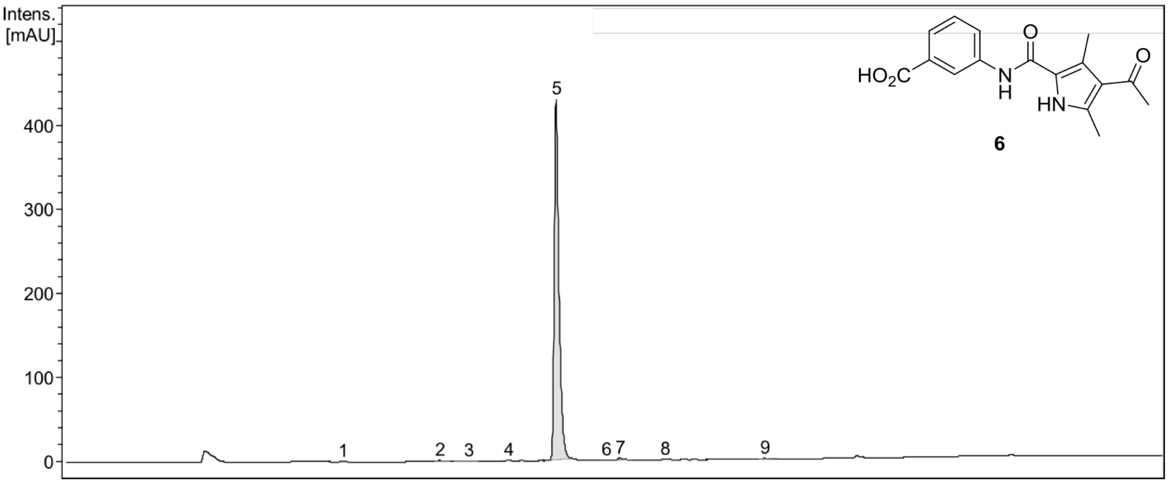
HPLC trace (for purity) of synthesized compounds



#	RT [min]	Area	Area Frac. %
1	0.9	3.4885	0.67
2	1.3	0.5209	0.10
3	1.4	0.3517	0.07
4	2.1	12.0190	2.3
5	2.6	504.3851	96.34
6	3.1	2.7686	0.53



#	RT [min]	Area	Area Frac. %
1	1.5	4.2666	0.56
2	1.8	0.7413	0.10
3	1.9	1.6912	0.22
4	2.0	2.3257	0.30
5	2.2	753.8320	98.76
6	2.4	0.4699	0.06



#	RT [min]	Area	Area Frac. %
1	1.3	0.3230	0.05
2	1.7	1.4703	0.24
3	1.8	1.6665	0.28
4	1.9	2.1705	0.36
5	2.1	588.4345	97.92
6	2.3	0.6963	0.12
7	2.4	4.0236	0.67
8	2.6	1.0989	0.18
9	3.0	1.0456	0.17

Bibliography:

1. Dawson, M.A., et al., *Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia.* Nature, 2011. **478**(7370): p. 529-533.
2. Bamborough, P., et al., *Fragment-based discovery of bromodomain inhibitors part 2: optimization of phenylisoxazole sulfonamides.* J Med Chem, 2012. **55**(2): p. 587-596.
3. Hewings, D.S., et al., *3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands.* Journal of Medicinal Chemistry, 2011. **54**(19): p. 6761-6770.
4. Borah, J.C., et al., *A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes.* Chem Biol, 2011. **18**(4): p. 531-541.
5. Kabsch, W., *Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants.* J. Appl. Crystallogr, 1993. **26**: p. 795-800.
6. Bailey, S., *The CCP4 Suite: Programs for Protein Crystallography.* Acta Crystallographica Section D-Biological Crystallography, 1994(50): p. 760-763.
7. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
8. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination.* Acta Crystallographica Section D-Biological Crystallography, 2002. **58**: p. 1948-1954.

Chapter 5

Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain

Unzue, A., Xu, M., Dong, J., Wiedmer, L., Spiliotopoulos, D., Caflisch, A., Nevado, C., *Journal of Medicinal Chemistry 2015.*

Journal of
Medicinal
Chemistry

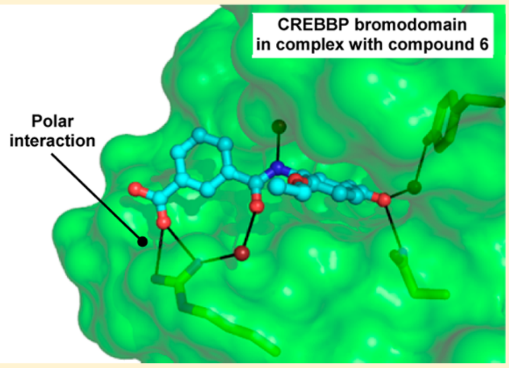
Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain

Andrea Unzue,[†] Min Xu,[‡] Jing Dong,[‡] Lars Wiedmer,[‡] Dimitrios Spiliotopoulos,[‡] Amedeo Caflisch,*[,‡] and Cristina Nevado*[,†]

[†]Department of of Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland
[‡]Department of of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland

Supporting Information

ABSTRACT: Novel ligands of the CREBBP bromodomain were identified by fragment-based docking. The in silico discovered hits have been optimized by chemical synthesis into selective nanomolar compounds, thereby preserving the ligand efficiency. The selectivity for the CREBBP bromodomain over other human bromodomain subfamilies has achieved by a benzoate moiety which was predicted by docking to be involved in favorable electrostatic interactions with the Arg1173 side chain, a prediction that could be verified a posteriori by the high-resolution crystal structure of the CREBBP bromodomain in complex with ligand **6** and also by MD simulations (see Xu, M.; Unzue, A.; Dong, J.; Spiliotopoulos, D.; Nevado, C.; Caflisch, A. Discovery of CREBBP bromodomain inhibitors by high-throughput docking and hit optimization guided by molecular dynamics. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b00171).



INTRODUCTION

The quest for bromodomain inhibitors as potential therapeutic tools has bloomed in recent years based on the growing understanding of epigenetic processes. Post-transcriptional modifications of histone tails constitute a highly sophisticated mechanism for gene expression control,[1,2] in which bromodomain proteins function as readers of the so-called histone code.[3−10] Given the direct connection between the regulation of gene expression and physiological and pathological processes, molecules interfering with bromodomains have recently emerged as chemical probes and/or clinical tools to regulate cancer, inflammation, and other diseases. Of the 61 human bromodomains known, the BET (bromo and extraterminal) family of bromodomains seems the most druggable,[4,10] as several potent ligands have been reported, in particular for BRD4(1) (the first bromodomain of the BRD4 protein).[11,12] These bromodomains have been directly connected to inflammation, aggressive types of squamous cell carcinomas and hematological malignancies such as acute myeloid leukemia.[3,6,12−15] The identification of potent and selective inhibitors would be extremely useful for other bromodomains, whose direct connection to specific diseases has not yet been established in many cases. The single bromodomain in the CREB binding protein (CREBBP) is a representative example, as few small molecule ligands of this bromodomain have been reported, and high selectivity proved to be difficult to achieve.[16−19] In 2014, Rooney et al. reported the first nanomolar potent CREBBP inhibitors based on dihydroquinoxalinone scaffolds.[20] Absolute stereocontrol was required to attain the desired potency, and still only moderate selectivity against BRD4(1), the most promiscuous bromodomain, was found. Last year, Hay et al. reported a medicinal chemistry optimization campaign for selective CREBBP bromodomain inhibitors starting from a nonselective 3,5-dimethylisoxazole ligand.[21]

We decided to take a different approach based on high-throughput docking on the target[22,23] and started our CREBBP ligand-identification campaign by fragment-based docking into the structure of the CREBBP bromodomain. Herein we present the result of a computer-aided, structure-based approach which has enabled the discovery of several nanomolar ligands of the CREBBP bromodomain upon optimization of binding by modulation of the electrostatic interaction with Arg1173 (numbering from PDB structures 3SVH and 3P1C), a residue located at the entrance of the binding site that is considered to be key for attaining selectivity toward CREBBP.[20,21,24−26]

RESULTS AND DISCUSSION

Computation. Details about the docking campaign carried out to identify the initial bromodomain ligand hits can be found in ref 27. Initially, 17 compounds (cluster representatives) were chosen for in vitro validation. The acetyl benzene **A** showed activity and was selected for further investigations based on its amenability to subsequent chemical editing. Three additional acetylbenzene derivatives (compounds **B**, **C**, and **D**) were also

Special Issue: Epigenetics

Received: January 29, 2015

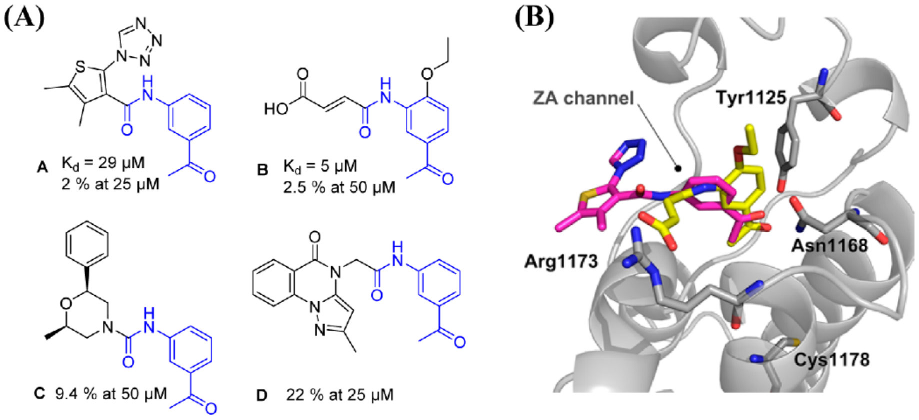
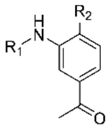
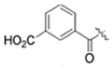
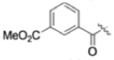
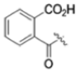


Figure 1. (A) Commercially available compound **A** identified in the high-throughput docking campaign and its available analogues **B**−**D**. Equilibrium dissociation constant (K_d) values and/or single-dose data (where lower percentages indicate stronger hits)[30] were determined in the competition binding assay (BROMOscan at DiscoveRx[28,29]) . The common structural features of **A**−**D** are shown in blue. (B) Docked poses of **A** (carbon atoms in magenta) and **B** (carbon atoms in yellow) into the CREBBP bromodomain (gray). The side chains of the conserved Tyr1125 and Asn1168 residues together with Arg1173 and Cys1178 are shown as sticks. The ZA channel, a region of the binding site that has proven to be important for bromodomain ligand design, is also indicated.[3,12,16]

Table 1. First Approach toward the Optimization of Compound B



Cmpd	R$_1$	R$_2$	ΔT$_m$ (°C)[a]	K$_d$ (µM)[b]		IC$_{50}$ (µM)[c]
B	Fumaric acid	OEt	1.8 ± 0.2	4.7	5.2	–
1	Fumaric acid	OMe	1.1 ± 0.1	–	–	60
2	Fumaric acid	Me	0.2 ± 0.1	–	–	>100
3	Fumaric acid	F	−0.1 ± 0.1	–	–	>100
4	Fumaric acid	H	0.1 ± 0.1	–	–	30
5	1,2-Cyclopropanedicarboxylic acid	OEt	1.1 ± 0.1	–	–	>100
6	 HO$_2$C	OEt	3.8 ± 0.1	0.88	0.67	8.7
7	 MeO$_2$C	OEt	2.8 ± 0.5[d]	0.69	0.49	5.0
8	 CO$_2$H	OEt	1.9 ± 0.1	–	–	64

[a]Median value of the shift in the melting temperature (number of measurements, >10) for the CREBBP bromodomain. SEM values are given. [b]K_d values were determined by a competition binding assay[28,29] in duplicate. Both recorded values are given in the table. [c]IC$_{50}$ values were determined at BPS Bioscience by means of a TR-FRET assay in duplicate. [d]The lower thermal shift of compound **7** in comparison to compound **6** could be due to partial hydrolysis of the methyl ester under the assay conditions.

purchased and tested (Figure 1A). The good affinity of compound **B** for CREBBP (K_d of 5 μM determined by a competition binding assay[28,29]) prompted us to focus our attention on the presence of the polar interactions formed between the molecule's carboxylic acid and the Arg1173 residue, which could be playing an important role in binding (Figure 1B). The acrylamide moiety in compound **B** could potentially establish a covalent bond with nucleophilic amino acid residues of the bromodomain, such as Cys1178. However, the formation of this bond is very unlikely because the sulfhydryl group of Cys1178 is completely buried (Figure 1B).

Optimization and Biophysical Characterization. Because of its potency, ligand efficiency (0.36 kcal/mol per heavy atom), and chemically novel blueprint within the CREBBP bromodomain inhibitors' pool,[31] we decided to focus our optimization campaign on compound **B**. Our strategy was to preserve its acetylbenzene moiety ("head group"), since the carbonyl group is predicted to be involved in a hydrogen bond with the conserved Asn1168 in the docked pose (Figure 1B).

Several modifications were thus designed along these lines in order to generate both potent and selective CREBBP inhibitors. First, we decided to keep the fumaric acid moiety and modulate the interactions of the substituents in ortho position relative to the amide group with the hydrophobic residues located on top of the binding site, Ile1122 and Leu1120. Replacement of the ethoxy group originally present in **B** with OMe, Me, or F substituents did not have a strong effect on affinity (Table 1, compounds **1**−**4**). We then turned our attention toward the side chain of the molecule while preserving the carboxylic acid moiety that interacts with Arg1173 according to the docked pose (Figure 1B). Remarkably, the substitution of the fumaric acid by an isophthalic group in **6** resulted in a 6-fold improvement in binding affinity representing the first nanomolar CREBBP inhibitor of our derivatization campaign. Devoid of a Michael acceptor system, **6** might be metabolically more stable preventing covalent binding to the protein through nucleophilic amino acid residues such as cysteins. We hypothesized that such improvement in affinity could be due to a more favorable interaction with the Arg1173 residue and the so-called LPF shelf together with a less unfavorable entropic penalty compared to the slightly more flexible fumaric acid derivative. The methyl isophthalate derivative **7** showed similar potency as **6**, indicating that its ester is involved in favorable polar interactions with Arg1173.

The crystal structure of compound **6** in complex with the CREBBP bromodomain (PDB code 4TQN, Figure 2) revealed an overall binding mode similar to the docked pose of compound **6** and the fumaric acid derivative **B** obtained via flexible docking (Figure 1B).[27] According to the X-ray structure, compound **6** binds in the acetyllysine pocket, with its acetyl substituent involved in a hydrogen bond with the side chain of the conserved Asn1168 and a water-bridged hydrogen bond with the Tyr1125 side chain hydroxyl. The acetylbenzene head, together with its ethoxy substituent, presents high shape complementarity with the binding site. The NH group of the amide linker forms a hydrogen bond with a water molecule whose oxygen atom is located at a distance of 3.0 Å from the carbonyl oxygen of Pro1110 (in the LPF shelf): the geometry for a water-bridged hydrogen bond, however, is not ideal because the water oxygen is not in the plane of the Pro1110 carbonyl. In contrast, the carbonyl group of the amide linker is involved in a favorable water-bridged hydrogen bond with the guanidinium of the Arg1173 side chain. Crucially, the aromaticity of the benzoic acid provides the optimal orientation for the carboxylic acid of compound **6** to form a very favorable polar interaction with the Arg1173 guanidinium of the CREBBP bromodomain. Previously, electrostatic interactions between the sulfonate group of ischemin and this arginine residue were observed.[25] Moreover, Conway and co-workers already reported a relatively weak ligand of CREBBP and BET bromodomains bearing a carboxylic acid that, even if not in direct contact with Arg1173, could form weak electrostatic interactions with this positively charged residue, resulting in a preferential binding toward CREBBP.[26]

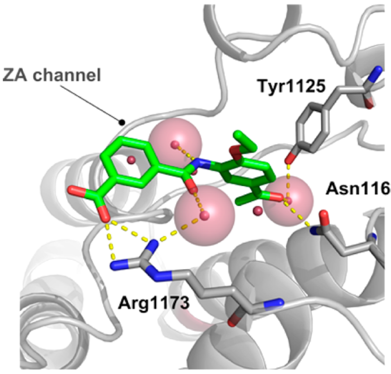


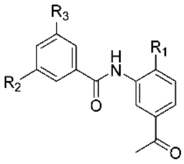
Figure 2. Crystal structure of CREBBP (gray) in complex with compound **6** (green) (PDB code 4TQN). The conserved Tyr1125 and Asn1168 residues together with Arg1173 are shown as sticks. The ZA channel, a region of the binding site that has proven to be important for bromodomain ligand design, is also indicated.[3,12,16]

Finite-difference Poisson calculations were performed with CHARMM[32,33] using the crystal structure of the CREBBP bromodomain in complex with compound **6** to evaluate the electrostatic contribution of the polar interaction to the free energy of binding of compounds **6** and **7**. The finite-difference Poisson calculation takes into account solvent screening effects which are significant because the Arg1173 side chain is partially exposed to the solvent (Figure 2). Despite the partial solvent accessibility, the electrostatic interaction between the carboxyl group of compound **6** and the guanidinium group of Arg1173 contributes −13.6 kcal/mol, which is about half of the total electrostatic interaction energy (−26.9 kcal/mol) between compound **6** and the CREBBP bromodomain. The finite-difference Poisson calculation was repeated on the energy minimized crystal structure. Again, the electrostatic interaction between the carboxyl group of compound **6** and the Arg1173 side chain guanidinium (−12.4 kcal/mol) is about half of the total electrostatic interaction energy (−25.0 kcal/mol). Analysis of the individual contributions to the binding free energy of compounds **6** and its methyl ester derivative **7** shows that the former has a slightly more favorable electrostatic contribution ($\Delta G^{6}_{\text{electr}} - \Delta G^{7}_{\text{electr}} = -0.8$ kcal/mol) whereas the latter has a more favorable van der Waals contribution ($\Delta G^{6}_{\text{vdW}} - \Delta G^{7}_{\text{vdW}} = 1.4$ kcal/mol) so that the difference in the calculated total binding free energy is close to zero, in line with the very similar dissociation constants of compounds **6** and **7** measured experimentally (Table 1).

With the MD simulation results[27] and high resolution crystal structure of the first nanomolar ligand **6** in hand we further explored the chemical space around this scaffold by introducing modifications at three different positions as illustrated in Table 2.

Thermal shift values similar to the one of compound **6** were observed for ligands bearing larger hydrophobic substituents in para position with respect to the acetyl group (**9**−**12**). More importantly, the replacement of the carboxylic acid by a tetrazole as isoster (**15**) maintained the affinity, whereas the presence of sulfonamides (**16**, **17**) resulted in a severe loss of activity. We then decided to enlarge our inhibitors toward the ZA channel by introducing both heteroaromatic (**18**, **19**) as well as linear substituents (**20**−**22**) bearing hydrogen bond acceptors or donors in R_3 that could interact with the amino acid residues located in the ZA loop. These modifications led to the most potent inhibitors of our optimization campaign, compounds **19**, **21**, and **22**, with K_d values of 170, 540, and 400 nM, respectively (Table 3). The similar thermal shift values measured with the CREBBP bromodomain and its paralogue, EP300 (Table 2) are consistent with the fact that identical residues are present in the acetyllysine binding site of both proteins. Compounds **23** and **24** were synthesized (Figure 3) to explore the effect of the amide direction (**23**) and the

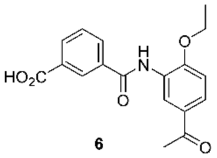
Table 2. Optimization of Compound 6[a]



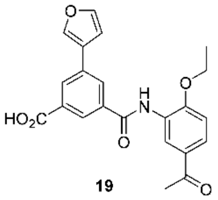
				ΔT_m (°C)[b]		
compd	R_1	R_2	R_3	CBP	EP300	IC_{50} (μM)[c]
6	OEt	CO_2H	H	3.8 ± 0.1	3.4 ± 0.2	8.7
9	OPr	CO_2H	H	3.8 ± 0.1	3.4 ± 0.2	7.4
10	O^{i}Bu	CO_2H	H	3.5 ± 0.1	3.0 ± 0.3	NA
11	OCH_2Cyc	CO_2H	H	3.4 ± 0.1	3.6 ± 0.4	NA
12	OBn	CO_2H	H	3.3 ± 0.1	2.6 ± 0.5	NA
13	Morph	CO_2H	H	1.4 ± 0.2	1.3 ± 0.2	NA
14	Cyc	CO_2H	H	3.1 ± 0.1	3.1 ± 0.2	15
15	OEt	1-tetrazole	H	3.3 ± 0.2	2.6 ± 0.4	7.5
16	OEt	$PhNHSO_2$	H	0.0 ± 0.1	0.0 ± 0.2	>10
17	OEt	$PhSO_2NH$	H	0.5 ± 0.1	0.4 ± 0.1	NA
18	OEt	CO_2H	4-pyridyl	4.1 ± 0.1	3.9 ± 0.2	3
19	OEt	CO_2H	3-furyl	5.2 ± 0.2	5.9 ± 0.2	1[d]
20	OEt	CO_2H	CH_2OH	3.8 ± 0.1	3.7 ± 0.3	3
21	OEt	CO_2H	CH_2OMe	5.1 ± 0.1	4.6 ± 0.3	2[d]
22	OEt	CO_2H	CH_2OTHP	6.0 ± 0.1	5.9 ± 0.2	1[d]

[a] Cyc, cyclopropyl; Morph, *N*-morpholyl; Bn, benzyl; THP, 2-tetrahydropyrane. [b] Median value of the shift in the melting temperature (number of measurements, >4) for CBP and EP300 proteins. SEM values are indicated. [c] IC_{50} values for the CREBBP bromodomain were determined by means of a TR-FRET assay in duplicate ("NA": an IC_{50} value could not be obtained at the tested concentration range). [d] K_d values for **19**, **21**, and **22** were determined by a competition binding assay[28,29] in duplicate with average values of 0.17, 0.54, and 0.40 μM, respectively.

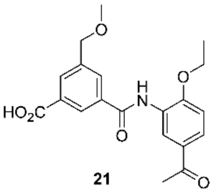
Table 3. Activity and Selectivity of the Most Potent Acetylbenzene Derivatives



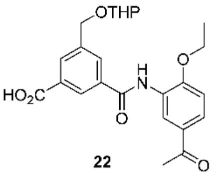
6



19



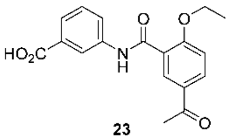
21



22

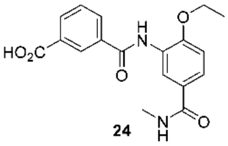
				K_d (μM) competition binding assay[b]									ΔT_m (°C)[c]		
compd	LE[a]	LLE	K_d (μM) ITC, CBP	CBP		BRD4(1)		BRD4(2)		BRD2(1,2)		S[d]	CBP	EP300	BRD4(1)
6	0.35	3.7	2.0	0.88	0.67	>50	>50	>50	>50	>50	>50	>65	3.8	3.4	0.4
19	0.32	3.2	0.3	0.17	0.17	10	9.8	>50	36	24	49	59	5.2	5.9	1.5
21	0.32	4.0	0.8	0.53	0.56	27	25	>50	>50	>50	>50	48	5.1	4.6	1.3
22	0.28	3.8		0.43	0.37								6.0	5.9	1.7

[a] LE = ligand efficiency, calculated as (ΔG/number of heavy atoms), is reported in kcal/mol per heavy atom; LLE = lipophilic ligand efficiency (calculated as pK_d − clogP).[34,35] clogP was calculated using ChemDraw. K_d values were determined by the competition binding assay BROMOscan and were considered when calculating LE and LLE values. [b] K_d values were determined by a competition binding assay[28,29] in duplicate. Both recorded values are given in the table. [c] Median value of the shift in the melting temperature (number of measurements, >9). SEM values did not exceed 0.3 °C. [d] Selectivity (*S*) between the CREBBP and BRD4(1) bromodomains determined by the ratio of K_d values obtained via the competition binding assay.[28,29]



23

ΔTm = 3.4 ± 0.1 °C
K_d = 2.5 μM



24

ΔTm = 1.1 ± 0.2 °C
K_d = 37 μM

Figure 3. Synthesized compounds **23** and **24**. The K_d values determined by a competition binding assay[28,29] and thermal shift values for the CREBBP bromodomain are indicated.

presence of an *N*-methylcarboxamide as a surrogate of the acetyl moiety (**24**),[31] respectively. In both cases a decrease in affinity for CREBBP was observed, confirming the excellent shape complementarity between the acetylbenzene moiety and the acetyllysine binding site of the CREBBP bromodomain and its paralogue EP300.

The selectivity evaluation of compounds **1**, **6**, **7**, **9**, **14**, **15**, and **19**−**23** as determined by thermal shift measurements (ΔT_m) against a panel of seven bromodomains is shown in Figure 4A. Remarkable target selectivity for CREBBP and EP300 over the other five human bromodomains (each belonging to a different subfamily) was found for the

DOI: 10.1021/acs.jmedchem.5b00172
J. Med. Chem. XXXX, XXX, XXX−XXX

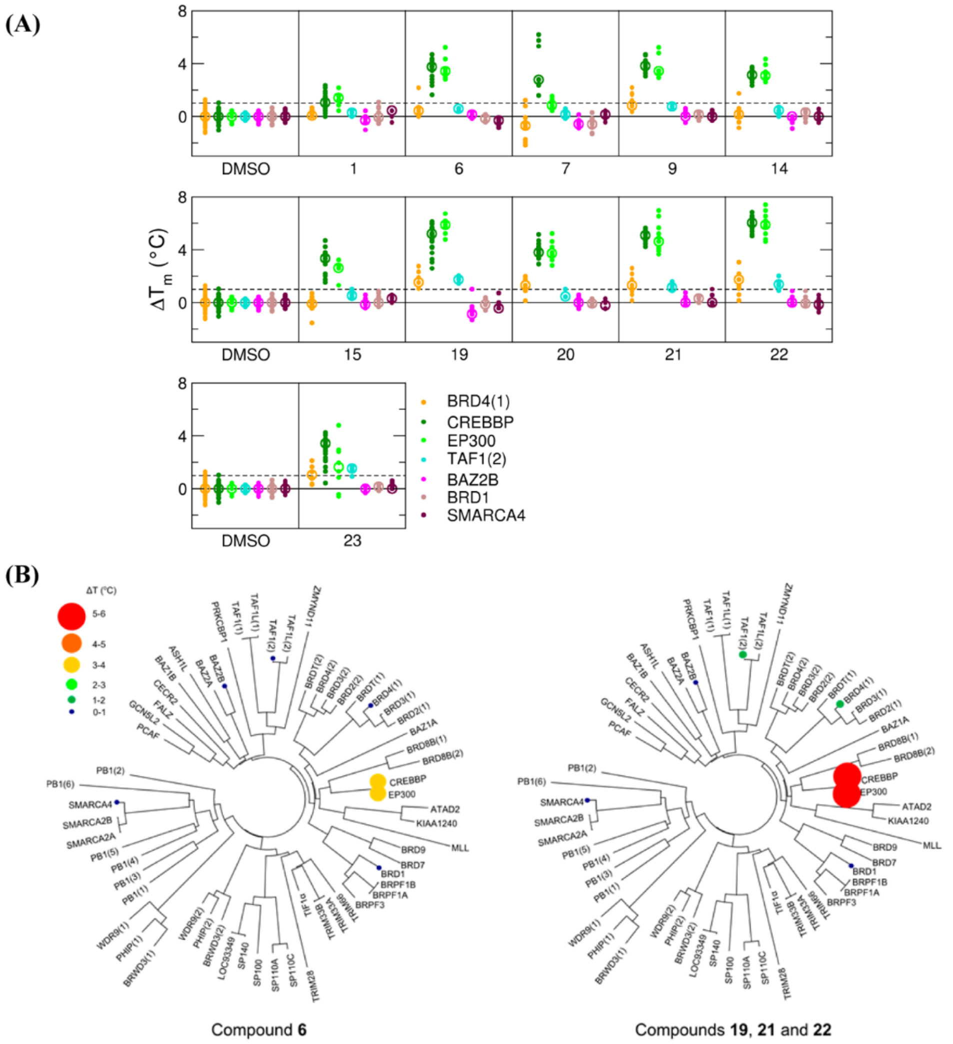


Figure 4. (A) Selectivity evaluation of compounds **1**, **6**, **7**, **9**, **14**, **15**, and **19**−**23** as determined by thermal shift measurements (ΔT_m) against a panel of seven bromodomains. Independent measurements are shown as dots, and the median is shown as a circle. The dashed line at 1 °C is an arbitrary threshold. (B) Bromodomain phylogenetic tree showing the selectivity evaluation of compounds **6** (left), **19**, **21**, and **22** (right) as determined by thermal shift measurements (ΔT_m) against a panel of different bromodomains. Sphere size and color indicate relative ΔT_m values according to the legend, which is the same for the two figures. Compounds **19**, **21**, and **22** share the same phylogenetic tree.

nanomolar ligands **6**, **19**, **21**, and **22** according to thermal shift (Figure 4B). Interestingly, the most potent inhibitor **19**, showed only 1.7 °C thermal shift against BRD4(1), which is the most promiscuous bromodomain.[20,21]

The selectivity was further investigated by isothermal titration calorimetry (ITC) and a competition binding assay (Table 3).[28,29] Concerning ITC, the weak signal obtained for BRD4(1) does not allow a reliable fit of the titration curve (see Figure S6 in the Supporting Information).

The BRD4(1)-K_d/CREBBP-K_d ratio determined by competition binding yields selectivity factors of >65, 59, and 48 for compounds **6**, **19**, and **21**, respectively. Such selectivity toward the CREBBP bromodomain supports the importance of the polar interactions between the carboxylic acids and the positively charged Arg1173 for the design of selective CREBBP ligands. The leadlike properties of these molecules were determined by computing the ligand efficiency (LE)[34,35] and ligand-lipophilicity efficiency per unit of potency (LLE). As shown in Table 3, the high ligand efficiency (LE) was preserved throughout the optimization campaign. Moreover, the LE values (0.28−0.35 kcal/mol per heavy atom) and the ligand-lipophilic efficiency (LLE) values (close to 4.0) of our nanomolar ligands **6**, **19**, **21**, and **22**, compare positively with those of previously disclosed ligands of the CREBBP bromodomain (see Table S5 in the Supporting Information).[18,20,21]

Screening of our most potent CREBBP inhibitors against a panel of 10 cancer cell lines showed lack of overall toxicity and significant growth inhibition for the leukemia cell lines MOLM-13, ML2, and HL-60 specifically for compounds **6**, **9**, **22**, and **23** (see Tables S3 and S4 in the Supporting Information). The methyl esters derived from compounds **6**, **9**−**11**, **19**, and **20**−**23** were also tested showing an equivalent or higher cell growth inhibitory activity compared to the corresponding acid counterparts, probably due to their better cell permeability. In fact, the methyl ester of **22** presented toxicity values in the low micromolar range, with GI_{50} values of 14.4 and 5.3 μM for ML2 and MOLM-13, respectively. Even if these results are still rather preliminary, the selectivity of compound **22** toward acute myeloid leukemia suggests a potential involvement of CREBBP in this pathology (see section 10 and Tables S3 and S4 of the

Supporting Information for further details on cell based experiments).[36−38]

■ CONCLUSIONS

In summary, we have discovered in silico and optimized by chemical synthesis a series of nanomolar potent and selective acetylbenzene ligands of the CREBBP bromodomain. Fragment-based, high-throughput docking was employed for the identification of novel scaffolds (see ref 27) whose affinity was enhanced in a straightforward manner through interactions within the ZA channel by introducing both heteroaromatic as well as linear substituents bearing hydrogen bond acceptors or donors that interact with the amino acid residues located in the ZA loop. The direct polar interaction between the benzoic acid moiety of these ligands and the Arg1173 guanidinium was exploited for the design of selective ligands of the CREBBP bromodomain. The potency, specificity, and easy synthetic availability of our compounds will be useful to unravel the role of CREBBP in several types of solid tumors and hematological malignancies.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

General procedures for the fragment docking, finite difference Poisson calculations, synthesis and characterization, biophysical and biological evaluation of final compounds, X-ray crystal structure refinement data, and a csv file of molecular formula strings. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.5b00172.

Accession Codes

PDB code for CREBBP in complex with the ligand **6** is 4TQN.

■ AUTHOR INFORMATION

Corresponding Authors

*A.C.: e-mail, caflisch@bioc.uzh.ch; phone, +41 44 635 55 21.
*C.N.: e-mail, cristina.nevado@chem.uzh.ch; phone, +41 44 635 39 45.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Lisa Gartenmann and Jean-Rémy Marchand are acknowledged for carrying out the thermal shift measurements and finite-difference Poisson calculations, respectively. We thank Lisa Caflisch and Dr. Ilian Jelezarov for support for protein purification and calorimetry, respectively. This work was supported by the Swiss Cancer Society (Krebsliga) and the Swiss National Science Foundation. D.S. is a recipient of the SystemsX.ch translational postdoc fellowship. We thank the Structural Genomics Consortium at Oxford University for providing the plasmids of all the bromodomains but EP300 (provided by AddGene).

■ ABBREVIATIONS USED

ALTA, anchor-based library tailoring approach; BET, bromodomain and extra terminal domain; Bn, benzyl; BRD2(1)/(2), first/second bromodomain of BRD2; BRD4-(1)/(2), first/second bromodomain of BRD4; CHARMM, chemistry at Harvard molecular mechanics; CREBBP/CBP, CREB binding protein; Cyc, cyclopropyl; DMSO, dimethyl sulfoxide; EP300, E1A binding protein p300; ^{i}Bu, isobutyl; ITC, isothermal titration calorimetry; Morph, *N*-morpholyl; Pr, propyl; SEED, solvation energy for exhaustive docking; SEM, standard error of the mean; THP, 2-tetrahydropyrane

■ REFERENCES

(1) Holliday, R. The Inheritance of Epigenetic Defects. *Science* **1987**, *238*, 163−170.

(2) Kouzarides, T. Chromatin modifications and their function. *Cell* **2007**, *128*, 693−705.

(3) Filippakopoulos, P.; Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discovery* **2014**, *13*, 337−356.

(4) Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* **2012**, *55*, 7346−7359.

(5) Prinjha, R. K.; Witherington, J.; Lee, K. Place your BETs: the therapeutic potential of bromodomains. *Trends Pharmacol. Sci.* **2012**, 33, 146−153.

(6) Müller, S.; Filippakopoulos, P.; Knapp, S. Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* **2011**, *13*, e29.

(7) Furdas, S. D.; Carlino, L.; Sippl, W.; Jung, M. Inhibition of bromodomain-mediated protein-protein interactions as a novel therapeutic strategy. *MedChemComm* **2012**, *3*, 123−134.

(8) Conway, S. J. Bromodomains: Are Readers Right for Epigenetic Therapy? *ACS Med. Chem. Lett.* **2012**, *3*, 691−694.

(9) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discovery* **2012**, *11*, 384−400.

(10) Zhang, G. T.; Sanchez, R.; Zhou, M. M. Scaling the Druggability Landscape of Human Bromodomains, a New Class of Drug Targets. *J. Med. Chem.* **2012**, *55*, 7342−7345.

(11) Müller, S.; Knapp, S. Discovery of BET bromodomain inhibitors and their role in target validation. *MedChemComm* **2014**, *5*, 288−296.

(12) Brand, M.; Measures, A. M.; Wilson, B. G.; Cortopassi, W. A.; Alexander, R.; Höss, M.; Hewings, D. S.; Rooney, T. P. C.; Paton, R. S.; Conway, S. J. Small Molecule Inhibitors of Bromodomain−Acetyl-lysine Interactions. *ACS Chem. Biol.* **2015**, *10*, 22−39.

(13) Lucas, X.; Wohlwend, D.; Hugle, M.; Schmidtkunz, K.; Gerhardt, S.; Schule, R.; Jung, M.; Einsle, O.; Gunther, S. 4-Acyl Pyrroles: Mimicking Acetylated Lysines in Histone Code Reading. *Angew. Chem., Int. Ed.* **2013**, *52*, 14055−14059.

(14) Mirguet, O.; Gosmini, R.; Toum, J.; Clement, C. A.; Barnathan, M.; Brusq, J. M.; Mordaunt, J. E.; Grimes, R. M.; Crowe, M.; Pineau, O.; Ajakane, M.; Daugan, A.; Jeffrey, P.; Cutler, L.; Haynes, A. C.; Smithers, N. N.; Chung, C. W.; Bamborough, P.; Uings, I. J.; Lewis, A.; Witherington, J.; Parr, N.; Prinjha, R. K.; Nicodeme, E. Discovery of Epigenetic Regulator I-BET762: Lead Optimization to Afford a Clinical Candidate Inhibitor of the BET Bromodomains. *J. Med. Chem.* **2013**, *56*, 7501−7515.

(15) Zhao, Y. J.; Yang, C. Y.; Wang, S. M. The Making of I-BET762, a BET Bromodomain Inhibitor Now in Clinical Development. *J. Med. Chem.* **2013**, *56*, 7498−7500.

(16) Hewings, D. S.; Rooney, T. P. C.; Jennings, L. E.; Hay, D. A.; Schofield, C. J.; Brennan, P. E.; Knapp, S.; Conway, S. J. Progress in the Development and Application of Small Molecule Inhibitors of Bromodomain-Acetyl-lysine Interactions. *J. Med. Chem.* **2012**, *55*, 9393−9413.

(17) Gerona-Navarro, G.; Yoel-Rodriguez; Mujtaba, S.; Frasca, A.; Patel, J.; Zeng, L.; Plotnikov, A. N.; Osman, R.; Zhou, M. M. Rational Design of Cyclic Peptide Modulators of the Transcriptional Coactivator CBP: A New Class of p53 Inhibitors. *J. Am. Chem. Soc.* **2011**, *133*, 2040−2043.

(18) Fedorov, O.; Lingard, H.; Wells, C.; Monteiro, O. P.; Picaud, S.; Keates, T.; Yapp, C.; Philpott, M.; Martin, S. J.; Felletar, I.; Marsden, B. D.; Filippakopoulos, P.; Müller, S.; Knapp, S.; Brennan, P. E. [1,2,4]Triazolo[4,3-a]phthalazines: Inhibitors of Diverse Bromodomains. *J. Med. Chem.* **2014**, *57*, 462−476.

(19) Hewings, D. S.; Fedorov, O.; Filippakopoulos, P.; Martin, S.; Picaud, S.; Tumber, A.; Wells, C.; Olcina, M. M.; Freeman, K.; Gill, A.; Ritchie, A. J.; Sheppard, D. W.; Russell, A. J.; Hammond, E. M.; Knapp, S.; Brennan, P. E.; Conway, S. J. Optimization of 3,5-Dimethylisoxazole Derivatives as Potent Bromodomain Ligands. *J. Med. Chem.* **2013**, *56*, 3217−3227.

(20) Rooney, T. P. C.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin, S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M. H.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Müller, S.; Knapp, S.; Brennan, P. E.; Conway, S. J. A Series of Potent CREBBP Bromodomain Ligands Reveals an Induced-Fit Pocket Stabilized by a Cation-pi Interaction. *Angew. Chem., Int. Ed.* **2014**, *53*, 6126−6130.

(21) Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P. C.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Müller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc.* **2014**, *136*, 9308−9319.

(22) Zhao, H. T.; Gartenmann, L.; Dong, J.; Spiliotopoulos, D.; Caflisch, A. Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2493−2496.

(23) Spiliotopoulos, D.; Caflisch, A. Molecular Dynamics Simulations of Bromodomains Reveal Binding-Site Flexibility and Multiple Binding Modes of the Natural Ligand Acetyl-Lysine. *Isr. J. Chem.* **2014**, *54*, 1084−1092.

(24) Sachchidanand; Resnick-Silverman, L.; Yan, S.; Mutjaba, S.; Liu, W. J.; Zeng, L.; Manfredi, J. J.; Zhou, M. M. Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* **2006**, *13*, 81−90.

(25) Borah, J. C.; Mujtaba, S.; Karakikes, I.; Zeng, L.; Müller, M.; Patel, J.; Moshkina, N.; Morohashi, K.; Zhang, W. J.; Gerona-Navarro, G.; Hajjar, R. J.; Zhou, M. M. A Small Molecule Binding to the Coactivator CREB-Binding Protein Blocks Apoptosis in Cardiomyocytes. *Chem. Biol.* **2011**, *18*, 531−541.

(26) Hewings, D. S.; Wang, M. H.; Philpott, M.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Vuppusetty, C.; Marsden, B.; Knapp, S.; Conway, S. J.; Heightman, T. D. 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J. Med. Chem.* **2011**, *54*, 6761−6770.

(27) Xu, M.; Unzue, A.; Dong, J.; Spiliotopoulos, D.; Nevado, C.; Caflisch, A. Discovery of CREBBP bromodomain inhibitors by high-throughput docking and hit optimization guided by molecular dynamics. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b00171.

(28) Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329−336.

(29) Quinn, E.; Wodicka, L.; Ciceri, P.; Pallares, G.; Pickle, E.; Torrey, A.; Floyd, M.; Hunt, J.; Treiber, D. Abstract 4238: BROMOscan - a high throughput, quantitative ligand binding platform identifies best-in-class bromodomain inhibitors from a screen of mature compounds targeting other protein classes. *Cancer Res.* **2013**, *73*, 4238.

(30) See Supporting Information for further details.

(31) Ferguson, F. M.; Fedorov, O.; Chaikuad, A.; Philpott, M.; Muniz, J. R. C.; Felletar, I.; von Delft, F.; Heightman, T.; Knapp, S.; Abell, C.; Ciulli, A. Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain. *J. Med. Chem.* **2013**, *56*, 10183−10187.

(32) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(33) Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(34) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.

(35) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881−890.

(36) Borrow, J.; Stanton, V. P.; Andresen, J. M.; Becher, R.; Behm, F. G.; Chaganti, R. S. K.; Civin, C. I.; Disteche, C.; Dube, I.; Frischauf, A. M.; Horsman, D.; Mitelman, F.; Volinia, S.; Watmore, A. E.; Housman, D. E. The translocation t(8;16)(p11, p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB binding protein. *Nat. Genet.* **1996**, *14*, 33−41.

(37) Sobulo, O. M.; Borrow, J.; Tomek, R.; Reshmi, S.; Harden, A.; Schlegelberger, B.; Housman, D.; Doggett, N. A.; Rowley, J. D.; ZeleznikLe, N. J. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 8732−8737.

(38) Goodman, R. H.; Smolik, S. CBP/p300 in cell growth, transformation, and development. *Gene Dev.* **2000**, *14*, 1553−1577.

Supporting Information

Fragment-based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain

Andrea Unzue,[†] Min Xu,[‡] Jing Dong,[‡] Lars Wiedmer,[‡] Dimitrios Spiliotopoulos,[‡] Amedeo Caflisch[‡,*]
and Cristina Nevado[†,*]

[†] Department of Chemistry and [‡] Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland

Table of contents

*To whom correspondence should be addressed. Phone: (41) 446353945. Fax: (41) 446353948.

E-mail: cristina.nevado@chem.uzh.ch, caflisch@bioc.uzh.ch

† Dept. of Chemistry

‡ Dept. of Biochemistry

1. Fragment-based high-throughput docking

The ALTA (anchor-based library tailoring) procedure was used for fragment-based high-throughput docking (Figure S1). The details of the procedure are presented in the preceding paper.

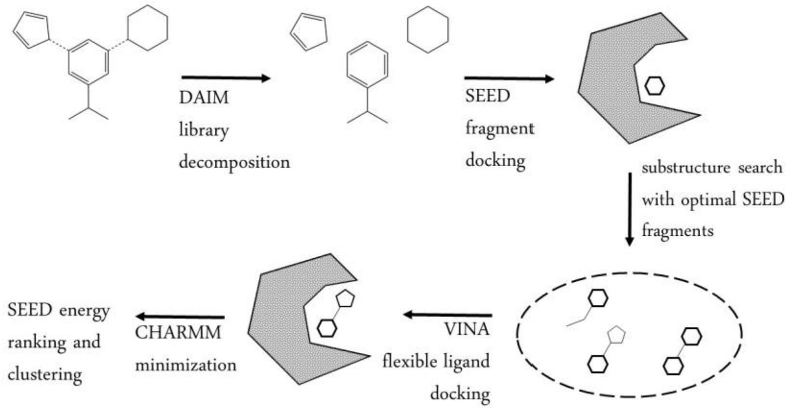


Figure S1. Application of the ALTA (anchor-based library tailoring) procedure[1-2] to the CREBBP bromodomain. The program DAIM automatically decomposes molecules into fragments by cutting at rotatable bonds.[3] CHARMM atom types and non-bonding parameters used in SEED[4-6] were generated by MATCH.[7] The program for fragment docking (called SEED[4-6]) requires about 5 seconds per fragment on a single core of an i7 CPU at 2.8 GHz. CHARMM minimization of the fragment with rigid CREBBP took about 2 seconds per fragment.

2. Finite-difference Poisson calculations

The electrostatic contribution to the binding free energy was evaluated by numerical solution of the Poisson equation using the finite-difference method as implemented in the *PBEQ* module[8] of the program CHARMM.[9] The solute/solvent dielectric discontinuity surface was delimited by the molecular surface spanned by the surface of a rolling probe of 1.4 Å. The dielectric constant of the solute and the solvent were set to 2.0 and 78.5, respectively. The six conserved water molecules were considered explicitly as part of the protein, i.e., they were assigned a dielectric constant of 2.0 as for the protein because they are essentially fixed in space and do not contribute to screening.The ionic strength was set to zero, and the temperature was set to 300 K. The size of the initial grid was determined by considering a layer of at least 20 Å around the solute. The partial charges of the solute were distributed on the grid points by the trilinear interpolation algorithm. First the linearized Poisson equation was solved on a grid of 1.0 Å spacing, which was followed by a focused calculation with a grid encompassing all of the solute and a grid spacing of 0.3 Å. For both calculations an iterative procedure (successive over-relaxation) was used. All calculations were carried out independently on the crystal structure (PDB code 4TQN) and the minimized crystal structure. The latter was obtained by conjugate gradient minimization with the CHARMM program (version 38b1), and the CHARMM param36 force field for CREBBP and the CHARMM general force field (which is compatible with CHARMM param36) for compound **6**.

4. Bromodomain expression and purification

Proteins were purified as described previously.[21] Briefly, His-tagged bromodomains were expressed in *Escherichia coli* BL21(DE3) cells upon induction with isopropyl thio-beta-D-galactoside (IPTG, final concentration 0.1 mM) for 16 h at 18 °C. Bacteria were lysated and (when required) the resulting extract was treated to remove DNA, adding 0.15% polyethylenimine (PEI). The His-tagged proteins were purified on HisTrap columns (GE Healthcare) and eluted using a step gradient of imidazole. The poly-Histidine tags were removed by overnight incubation with His-tagged tobacco etch virus (TEV) protease purified in-house (if required by the purification protocol, in the meantime the sample was exchanged via dialysis). A size-exclusion chromatography step (HiLoad 16/600 Superdex75 column) and a Ni-affinity chromatography step were subsequently performed to finally purify the cleaved bromodomains. Samples were then concentrated, flash frozen and stored at –80 °C.

5. X-ray crystallography

Crystallization, Data Collection, and Structure Determination

Crystals of the CREBBP bromodomain were grown at 4°C using the hanging drop vapor diffusion method. A 50 mM solution of compound **6** (in 100 % DMSO) was added into the CREBBP protein to reach a final DMSO concentration of 1 % (v/v) and the mixture was incubated on ice for 1 hour before crystallization. Then equal volumes of protein (with compound **6**) and reservoir solutions (0.1 M MES pH 6.5, 0.10 $MgCl_2$, 20 % PEG 6000, 10 % ethylene glycol) were mixed and crystals appeared after 1 to 2 days. The crystals were flash-frozen in liquid nitrogen with extra 10% ethylene glycol as cryoprotectant for measurements. Data sets were collected on a PILATUS 6MF detector at the Swiss Light Source beamline X06SA of the Paul Scherrer Institute (Villigen, Switzerland) and indexed, integrated and scaled with the XDS[22] and CCP4 programs.[23] The structures were solved by molecular replacement with PHASER[24] using the CREBBP structure (PDB entry 4NR5) as a search model and refined with PHENIX.[25] The atomic coordinates and structure factors of CREBBP in complex with inhibitor **6** have been deposited with the Protein Data Bank as entry 4TQN.

Table S1

	Compound **6**
Space group	P1 21 1
Unit cell	
a (Å)	24.94
b (Å)	42.94
c (Å)	51.98
alpha	90.00
beta	97.24
gamma	90.00
Resolution range (Å)	42.94 -1.70
Unique reflections	12119(1768)
<I/σ(I)>	15.6(5.3)
R merge	0.068(0.380)
Completeness (%)	99.9(99.2)
Multiplicity	6.5(6.2)
Refinement	
Resolution range (Å)	33.00-1.70
R factor/R free	0.1813/0.1990
Mean B factors (A2)	23.24
RMS bonds (Å)	0.006
RMS angles (°)	1.155

5.1 Composite Omit Map of Ligand 6

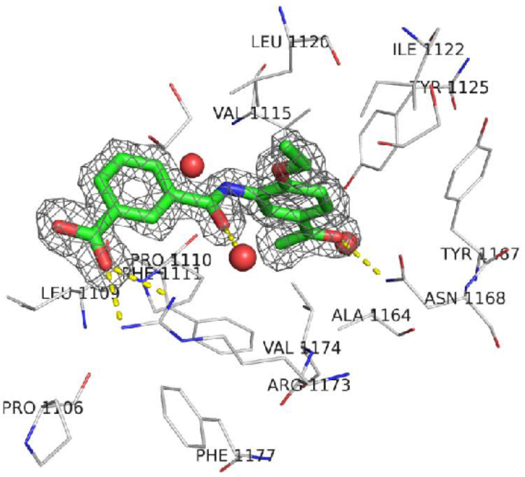


Figure S2. $2mF_o - DF_c$ electron density maps contoured at 1σ (grey mesh) were generated in a region within 1.6 Å for compound **6** using PHENIX and Pymol.

6. Thermal shift measurements

Thermal shift measurements were carried out as previously described[26] with a final volume of 20 µl, ligand and protein concentrations 100 µM (50 µM for compound **18**) and 2 µM, respectively. The reported values (ΔT_m) are calculated as the difference between the transition midpoints of an individual sample and the average of the reference wells (containing the protein and the DMSO only) in the same plate. DMSO concentration was kept at 0.2% (v/v).

6.1 Phylogenetic tree

The bromodomain sequence aligment previously reported by Filippakopoulos *et. al.* was used excluding the residues from the plasmid.[21] The evolutionary history was inferred using the Neighbor-Joining method.[27] The optimal tree with the sum of branch length = 21.33784094 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method[28] and are in the units of the number of amino acid substitutions per site. The analysis involved 61 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 57 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.[29]

6.2 Correlation between IC_{50} and Thermal Shift values with CREBBP

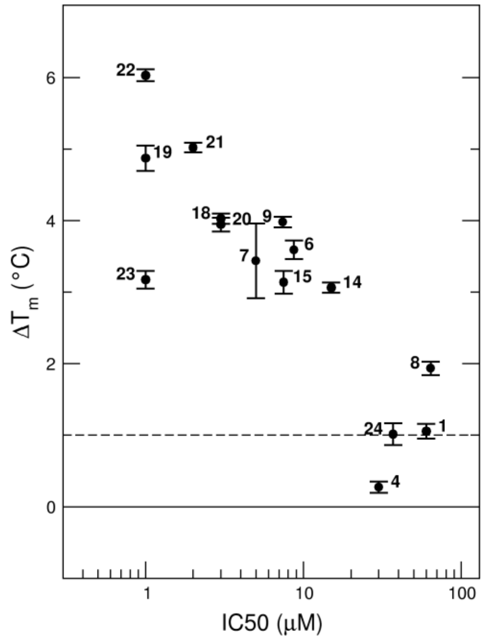


Figure S3. Correlation between IC_{50} values (determined by TR-FRET assay at BPS Bioscience) and thermal shift values (ΔT_m) obtained with CREBBP. Average and SEM of the thermal shift results are shown.

7. TR-FRET assays

TR-FRET assays were carried out in duplicate at BPS Bioscience using a recombinant CBP bromodomain (BPS catalogue #31128) and the BET Ligand (BPS catalogue #33000) as provided in the CREBBP TR-FRET Assay Kit (BPS catalogue #32619). A 10 mM solution of the compound under investigation in DMSO was prepared and shipped to BPS Bioscience, where it was tested at 10 concentrations over the range of 0.001-10 μM (compounds **6**, 7, **9-13**, **15-19**, **21-23**) or 0.01-100 μM (compounds **1-5**, **8**, **14**, **20**, **24**) . Each compound solution was then diluted in water to obtain a 10% DMSO solution. 2 μL of this dilution were added to a 20 μL reaction mixture (12.5 nM CBP, 125 nM BET Ligand, including FRET dyes and the amount of compound needed to reach the indicated concentration in the Table below). The resulting mixture was incubated for 2 hours at room temperature prior to reading the TR (time resolved)-FRET signal using a Tecan Infinite M1000 plate reader. The negative control consisted of the aforementioned mixture in which the buffer was added in place of compound. TR-FRET were recorded as the ratio of the fluorescence of the acceptor and the donor dyes (acceptor/donor).

The TR-FRET data was analyzed using Graphpad Prism software. The percent activity in the presence of each compound was calculated according to the following equation: % activity = [(F- Fb)/(Ft – Fb)]×100, where Ft is the TR-FRET signal in the absence of any compound (100 % activity), Fb the TR-FRET signal in the absence of the bromodomain (0 % activity) and F the TR-FRET signal in the presence of the compound. The percent inhibition was calculated according to the following equation: % inhibition = 100 - % activity. The values of % activity versus a series of compound concentrations were then plotted using non-linear regression analysis of Sigmoidal dose-response curve generated with the equation $Y=B+(T-B)/1+10^{((\mathrm{LogIC50}-X)\times \mathrm{Hill\ Slope})}$, where Y=percent activity, B=minimum percent activity, T=maximum percent activity, X= logarithm of compound and Hill Slope=slope factor or Hill coefficient. The IC_{50} value corresponds to the concentration causing a half-maximal percent activity.

8. BROMOscan assays

K_D and % binding of binding affinity determinations by means of BROMOscan technology was carried out at DiscoveRx. E. *coli* derived from BL21 strain was used as host to grow T7 phage strains displaying the bromodomains. E. *coli*, grown to log-phase, were infected with T7 phage (from a frozen stock, being the multiplicity of infection 0.4) and incubated while shaking at 32 °C for 90-150 minutes, until lysis. In order to remove cell debris, lysates were centrifuged at 5,000 x g and filtered (0.2 μm). Affinity resins were obtained by treating streptavidin-coated magnetic beads with biotinylated acetylated peptide ligands for 30 minutes at 25°C. Those beads were then blocked with excess of biotin and washed with blocking buffer (SeaBlock (Pierce), 1 % bovine serum albumin, BSA, 0.05 % Tween20, 1 mM dithiothreitol, (DTT) removing the unbound ligand and reducing non-specific phage binding.

During the experiment, the bromodomain, ligand-bound affinity beads and test compounds were combined in a buffer composed of 17% SeaBlock, 0.33x phosphate-buffered solution, PBS, 0.04% Tween20, 0.02% BSA, 0.004% sodium azide and 7.4 mM DTT. Test compounds were prepared as 50 mM in pure DMSO and diluted to 5 mM with monoethylene gycol, MEG (100× concentrated in respect to the top screening concentration, 50 μM). During the assay a DMSO and MEG final concentration of 0.1% and 0.9% respectively was used. The assays were carried out in polystyrene 96-well plates in a final volume of 0.135 mL. The assay plates were incubated at 25 °C with shaking for 1 hour and the affinity beads were washed with a buffer composed of 0.05% Tween 20 in PBS. The beads were then re-suspended in the elution buffer (1x PBS, 0.05% Tween 20, 2 μM non-biotinylated affinity ligand) and incubated at 25°C with shaking for 30 minutes. The bromodomain concentration in the eluates was measured by qPCR. Binding constants (K_d) were calculated with a standard dose-response curve using the Hill equation and curves were fitted using a non-linear least square fit with the Levenberg-Marquardt algorithm. The % of binding interactions are reported as '% Ctrl', where lower numbers indicate stronger hits. The corresponding values were obtained as follows:

$$\% \, ctrl = \frac{test\ compound\ signal - positive\ control\ signal}{negative\ control\ signal - positive\ control\ signal} \times 100$$

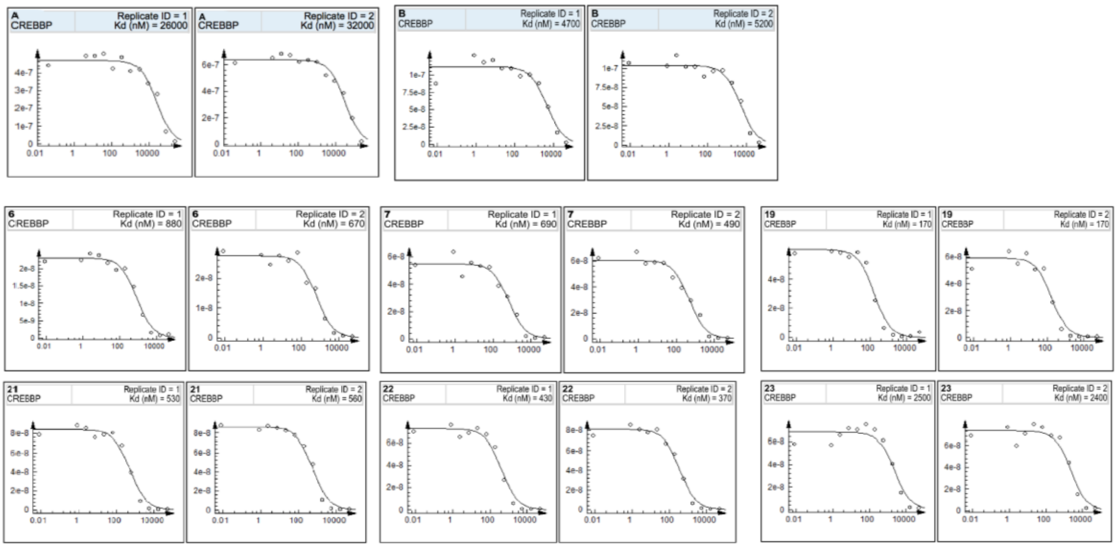


Figure S4. Dose response curves for the binding of the CREBBP bromodomain to compounds **A**, **B**, **6**, **7**, **19** and **21**-**23** performed at DiscoveRx.

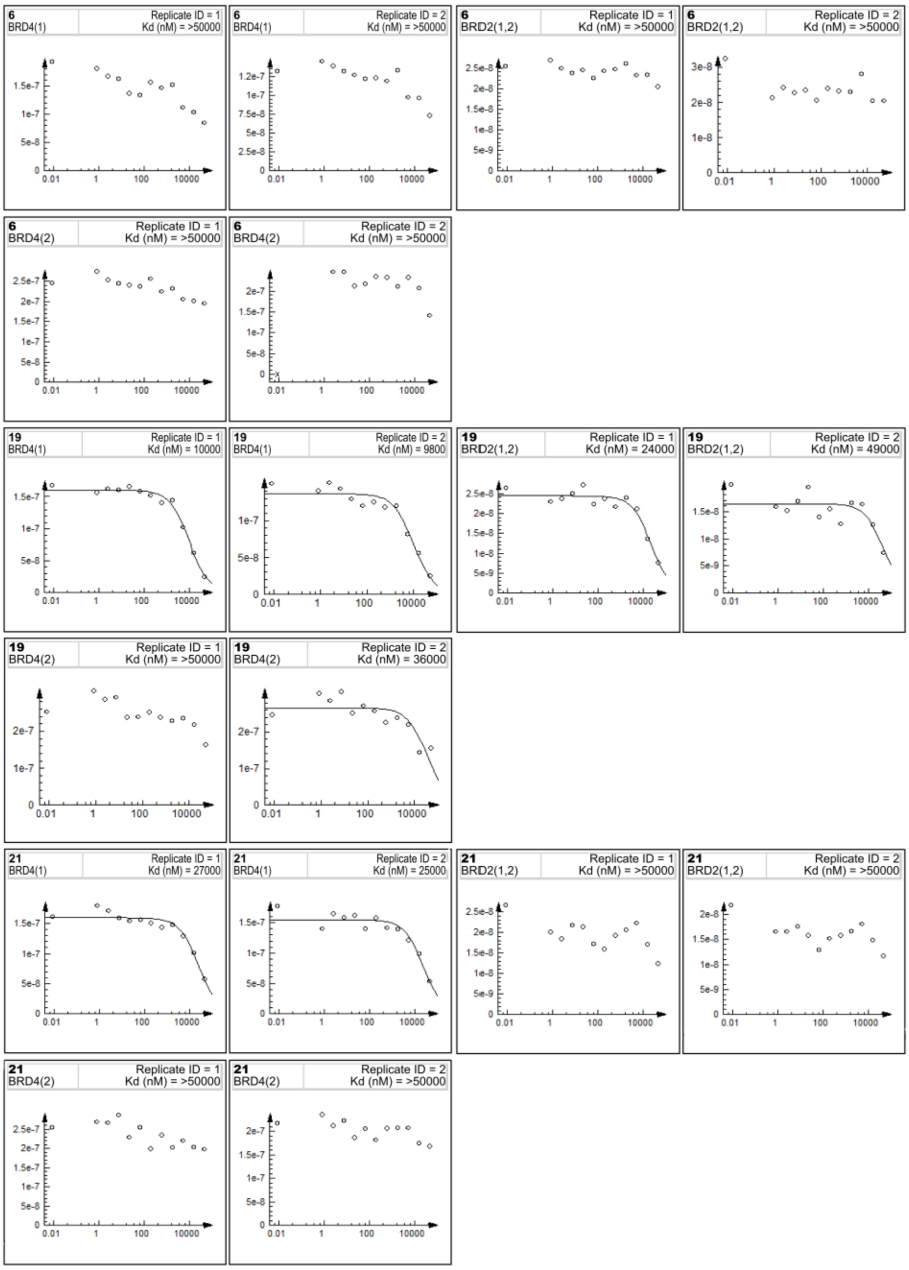


Figure S5. Dose response curves for the binding of the BET bromodomains to compounds **6**, **19** and **21** performed at DiscoveRx.

9. ITC EXPERIMENTS

Isothermal Titration Calorimetry experiments were performed on a VP-ITC instrument (MicroCal, Inc., Northampton, MA). Protein samples thoroughly dialyzed against the same batch of buffer in order to minimize artifacts due to minor differences in buffer composition and the protein concentration was determined after a filtering through a 0.22 µm pore-size filter.

Bromodomains (300-500 µM) were injected into the 1.4-mL sample cell containing the compound (50 µM) dissolved into the ITC buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1% DMSO). The titration experiments were carried out at 15°C while stirring at 300 rpm: after a control injection of 2 µL, 29

10-μL injections (10 s duration, with a 4 min interval between) were performed. The raw data were integrated, normalized for concentration, and analyzed using a single binding site model, supplied with the MicroCal Origin software package to obtain the apparent K_d values (K_d^{app}). The actual K_d values were calculated assuming that DMSO acted as a competing ligand[30] according to the equation:[31]

$$K_d = \frac{K_d^{app}}{1 + \frac{[DMSO]}{K_d^{DMSO}}}$$

where [DMSO] and K_d^{DMSO} indicate the DMSO concentration in the ITC buffer and the dissociation constant of the DMSO for the bromodomain under investigation, respectively (Figure S6).

The K_d^{DMSO} value for CREBBP was calculated titrating the histone H3K56ac peptide (ac-IRRYQ(Kac)STELLY-am, where ac-, -am and Kac, indicate acetylation, amidation and acetylated lysine side chain, respectively, purchased at GenScript) into a CREBBP bromodomain solution in the presence and absence of 0.75% DMSO with protein and peptide concentrations of 688 or 740 μM and 85 μM, respectively (Figure S7).

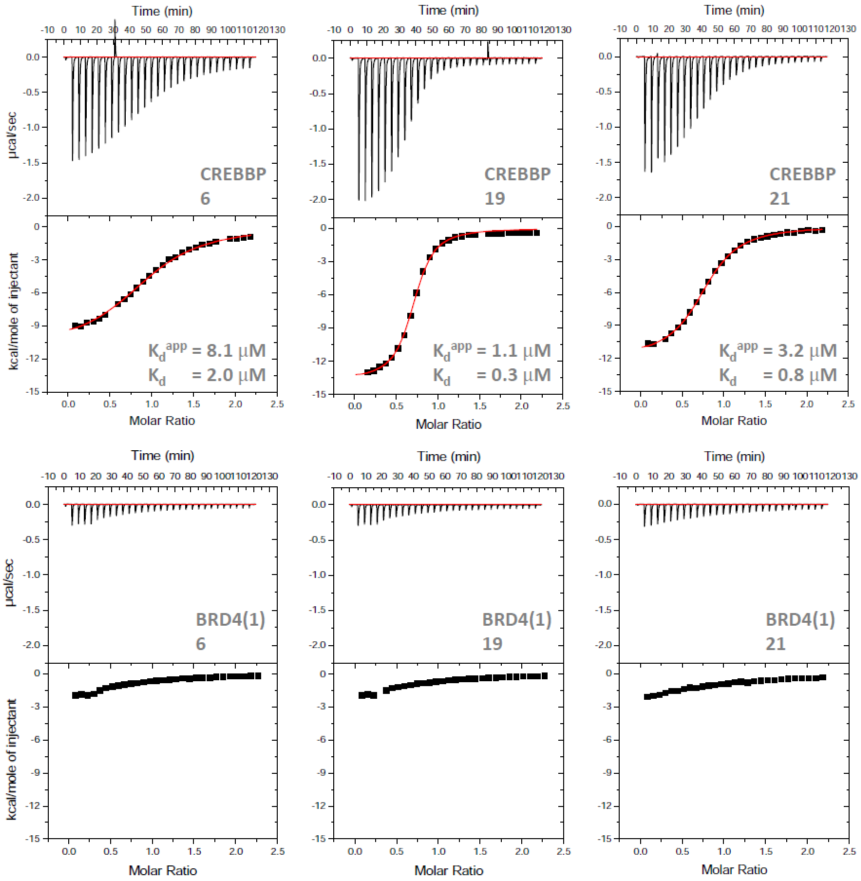
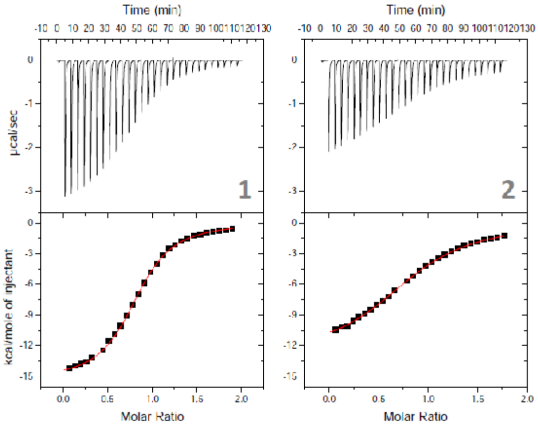


Figure S6. ITC titration curves for the binding of the CREBBP and BRD4(1) bromodomains (upper and lower panels, respectively) to compounds **6**, **19** and **21**.



exp	[DMSO] (M)	app K_d (µM)
1	0.000	4.77
2	0.106	15.71

$$K_{d_{cmpd}} = \frac{K_{d_{cmpd}}^{app}}{1 + \frac{[DMSO]}{K_{d_{DMSO}}}}$$

$$K_{d_{DMSO}} = 46.3\ mM$$

Figure S7. The N-terminally acetylated, C-terminaly amidated IRRYQ(Kac)STELLY peptide was titrated into a CREBBP bromodomain solution in the absence of DMSO (experiment 1) and in the presence of 0.75% DMSO (experiment 2). The K_d for DMSO was calculated as shown in the right panel.

Table S2. Thermodynamic parameters measured by ITC for the CREBBP bromodomain. ΔG and ΔH values are given in kcal/mol, ΔS values are given in cal/mol/T.

Cpd	N	K_d^{app} (μM)	ΔG	ΔH	ΔS
6	0.99 ± 0.01	8.13 ± 0.43	–6.7	–10.9 ± 0.1	–14.5
19	0.70 ± 0.00	1.11 ± 0.08	–7.9	–13.7 ± 0.1	–20.1
21	0.79 ± 0.00	3.25 ± 0.10	–7.2	–11.9 ± 0.0	–16.2

10. Cell culture and cytotoxicity measurements

MDA-MB-231, HT-29 cells (obtained from the UZH Cancer Institute) and HeLa cells (obtained from Dr. Nathan Luedtke, Chemistry Department, UZH) were cultured in DMEM supplemented with 10 % (v/v) fetal bovine serum. K562 cells (obtained from Dr. Silvio Hemmi, Institute of Molecular Life Sciences, UZH), HL-60, ML2 (obtained from Dr. Nathan Luedtke, Chemistry Department, UZH) and HOP-92 (purchased from the NCI) were cultured using RPMI medium supplemented with 10 % (v/v) fetal bovine serum. Finally, AML3, PL-21 and MOLM-13 (obtained from Dr. Nathan Luedtke, Chemistry Department, UZH) were cultured using RPMI medium supplemented with 20 % (v/v) fetal bovine serum. All the media were additionally supplemented with 100 units/mL of penicillin, 100 μg/mL of streptomycin, 4.5 g/L glucose, 0.11g/L sodium pyruvate and 2mM glutamine and the cells were grown at 37 °C in 5 % CO_2 atmosphere with 80 % relative humidity.

MDA-MB-231, HT-29, HeLa, HOP-92 and PC-3 cells were plated at 10,000 cells per well (100 μL per well) in 96-well culture dishes and allowed to incubate for 24 h. The old media was removed, cells were washed with PBS (phosphate-buffered saline) and fresh medium was added. A 5 mM solution of inhibitor (in 100% DMSO) was serially diluted in the culture media (8 different concentrations were used) and allowed to incubate for 72 h (MDA-MD-231, HT-29, HOP-92, HL-60, ML2, AML3, PL-21, MOLM-13 and PC-3) or 48h (HeLa and K562). Control cells were treated with the same DMSO concentrations. After the incubation period the medium was removed, cells were washed with PBS to be then incubated with fresh medium containing 86 nM resazurin. Resazurin is reduced to the fluorescent resorufin in the mitochondria: the fluorescence intensity upon incubation in the presence of a living cell culture thus directly correlates with the metabolic viability of the cells. Fluorescence was quantified after 4 hours using a fluorescence microplate reader (Biotek, FLx800^{TM}, excitation and emission wavelengths 560 and 590 nm, respectively). The measured fluorescence values were corrected from the control samples containing DMSO.

Leukemia cell lines (K562, HL-60, ML2, AML3, PL-21 and MOLM-13) were seeded at a density of 20,000 cells per well in 100 μL of RMPI media in 96 well microtiter plates. After 24 hours, 12.5 μL of a 10 fold concentrated drug (or DMSO solution for the control) in RMPI media was added in every well. After 48 hour incubation, resazurin was added to every well to obtain a final concentration of 86 nM and, after 3 hours, cell viability was assessed by measuring the ability of the cells to process resazurin by quantifying the fluorescence using a fluorescence microplate reader (Biotek, FLx800™) as described above.

Table S3. Resazurin reduction (percentage of the control) upon incubation of compounds **6**, **9**, **10**, **15**, **19-23** for 48 or 72 h at a concentration of 50 μM in eleven different cancer cell lines. This preliminary screening of toxicity was done as a single experiment. NM indicates no metabolic activity change in comparison to the DMSO-treated cells.

	MDA-MD-231	**HT-29**	**HeLa**	**HOP-92**	**K562**	**HL-60**	**ML2**	**AML3**	**PL-21**	**MOLM-13**	**PC-3**
6	NM	NM	NM	80.3	98.3	77.1	98.3	86.8	90.2	65.2	87.7
9	90.7	NM	94.8	91.0	NM	62.8	NM	87.3	74.7	15.1	84.3
10	NM	NM	97.6	71.9	NM	87.8	97.3	90.0	86.0	NM	89.1
15	86.7	NM	93.2	97.8	81.3	43.2	64.1	63.0	59.4	50.2	94.1
19	NM	NM	84.5	65.3	84.8	42.1	53.2	68.9	69.2	49.4	65.0
20	84.3	NM	94.2	72.8	NM	90.3	NM	98.7	NM	76.3	99.9
21	94.4	NM	90.4	77.3	NM	84.7	NM	96.9	94.3	73.2	81.8
22	88.2	NM	89.3	78.3	90.8	45.7	26.5	69.9	70.5	43.0	64.1
23	NM	NM	NM	NM	99.2	68.6	NM	84.2	86.3	74.8	69.5

Table S4. Metabolic activity values (%) on three different leukemia cell lines at 50 μM compound concentration.[a]

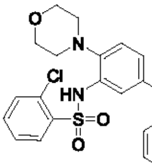
Cmpd	HL-60		ML2		MOLM-13	
	Acid	Ester	Acid	Ester	Acid	Ester
6	80	78	93	75	71	46 (46 μM)[b]
9	78	89	66	97	19 (33 μM)[b]	82
10	89	89	100	86	89	88
11	77	85	88	99	73	76
15	59	-	107	-	57	-
19	59	96	91	85	60	90
20	89	-	102	-	91	-
21	81	74	104	73	85	66
22	65	73	93	36 (14.4 μM)[b]	61	11 (5.3 μM)[b]
23	82	88	66	60	60	26 (37 μM)[b]

[a] Metabolic activity values were determined using resazurin reduction after 72 h incubation with the corresponding compound in comparison to DMSO treated cells at eight different concentrations (100-0.8 μM). The % of metabolic activity at 50 μM compound concentration is shown. The % of metabolic activity is given as the mean of at least three independent experiments. Variability around

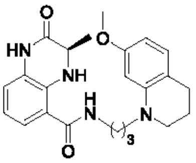
the mean value was < 26 % in all cases. [b] Values in parentheses are GI_{50} values as determined by triplicate experiments with less than 30 % standard error.

11. Comparison table to known CREBBP ligands

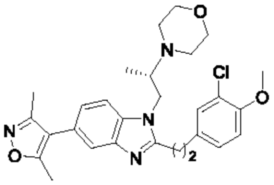
Table S5. Activity and selectivity of acyl benzene derivatives and comparison with previously reported CREBBP bromodomain ligands.



25



26



27

Cmpd	LE[a]	LLE	K_d (μM) ITC	K_d (μM) competition binding assay[32-33]					ΔT_m (°C)[b]		
			CBP	CBP	BRD4(1)	BRD4(2)	BRD2(1,2)	S[c]	CBP	EP300	BRD4(1)
This study											
6	0.35	3.7	2.0	0.77	>50	>50	>50	>65	3.6	3.4	0.5
19	0.32	3.2	0.3	0.17	10	36	36	59	4.9	5.9	1.7
21	0.32	4.0	0.8	0.54	26	>50	>50	48	5.0	4.6	1.3
22	0.28	3.8	–	0.40	-	-	-	-	6.0	5.9	1.6
Reported by others (main text, Figure 3, bottom)											
25[34]	0.25	2.4	–	0.20[d]	0.16[d]	-	-	0.8	7.6	–	4.4
26[35]	0.29	4.3	0.3	–	1.38[e]	-	-	4	5.4	–	–
27[36]	0.27	2.0	0.021	0.080[d]	0.85	5.2	-	40	9.7	9.7	1.8

[a] LE = ligand efficiency calculated as (ΔG/number of heavy atoms) is reported in kcal/mol per heavy atom; LLE = lipophilic ligand efficiency (calculated as pK_d-*c*logP),[37-38] clogP was calculated using ChemDraw. [b] Median value of the shift in the melting temperature (number of measurements between 9 and 38 per compound/protein pair). The largest SEM was 0.5 °C and most SEM values were below 0.2 °C. [c] Selectivity (S) between the CREBBP and BRD4(1) bromodomains determined by the ratio of K_d values obtained via the competition binding assay. It was not possible to calculate the selectivity using the K_d values obtained via ITC due to the impossibility to reliably fit the titration curves for the BRD4(1) bromodomain. [d] Potency determined by AlphaScreen. [e] Potency determined by Isothermal Titration Calorimetry (ITC).

References

1. Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: Discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73*, 11-18.

2. Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a Novel Chemotype of Tyrosine Kinase Inhibitors by Fragment-Based Docking and Molecular Dynamics. *ACS Med. Chem. Lett.* **2012**, *3*, 834-838.

3. Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking. *J. Med. Chem.* **2006**, *49*, 7384-7392.

4. Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001**, *42*, 256-268.

5. Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098-8106.

6. Scarsi, M.; Apostolakis, J.; Caflisch, A. Comparison of a GB Solvation Model with Explicit Solvent Simulations: Potentials of Mean Force and Conformational Preferences of Alanine Dipeptide and 1,2-Dichloroethane. *J. Phys. Chem. B* **1998**, *102*, 3637-3641.

7. Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L. MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *J. Comput. Chem.* **2012**, *33*, 189-202.

8. Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59-75.

9. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614.

10. Baldwin, J. E.; Cianciosi, S. J. Syntheses of Racemic and Both Chiral Forms of Cyclopropane-1,2-D(2) and Cyclopropane-1-C-13-1,2,3-D(3). *J. Am. Chem. Soc.* **1992**, *114*, 9401-9408.

11. Benatti, L.; Fariello, R.; Salvati, P.; Pellicciari, R.; Caccia, C. *Newron Pharm. Spa.* **2004**, *EP1424333 (A1)*.

12. Voitekhovich, S. V.; Vorobiov, A. N.; Gaponik, P. N.; Ivashkevich, O. A. Synthesis of new functionally substituted 1-R-tetrazoles and their 5-amino derivatives. *Khim Geterotsikl+* **2005**, 1174-1179.

13. Bridger, G.; Skerlj, R.; Kaller, A.; Harwig, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, S.; Smith, C. D.; Fluri, M. R. D. Chemokine receptor binding heterocyclic compounds **2002**, *US2002147192 (A1)*.

14. Wydysh, E. A.; Medghalchi, S. M.; Vadlamudi, A.; Townsend, C. A. Design and Synthesis of Small Molecule Glycerol 3-Phosphate Acyltransferase Inhibitors. *J. Med. Chem.* **2009**, *52*, 3317-3327.

15. Feutrill, J.; Leriche, C.; Middlemiss, D. *Fovea Pharmaceuticals* **2013**, *WO2013037705 (A2)*.

16. Meng, L.; Cheng, Q. G.; Kim, C.; Gao, W. Y.; Wojtas, L.; Chen, Y. S.; Zaworotko, M. J.; Zhang, X. P.; Ma, S. Q. Crystal Engineering of a Microporous, Catalytically Active fcu Topology MOF Using a Custom-Designed Metalloporphyrin Linker. *Angew. Chem. Int. Ed.* **2012**, *51*, 10082-10085.

17. Zhao, G. Z.; Chen, L. J.; Wang, W.; Zhang, J.; Yang, G.; Wang, D. X.; Yu, Y.; Yang, H. B. Stimuli-Responsive Supramolecular Gels through Hierarchical Self-Assembly of Discrete Rhomboidal Metallacycles (vol 19, pg 10094, 2013). *Chem. Eur. J.* **2013**, *19*, 12923-12923.

18. Singh, R.; Zhou, H. Mass tags for quantitative analysis **2002**, *US2002164649 (A1)*.

19. Pamukcu, R.; Piazza, G. A. Method of treating a patient having precancerous lesions with phenyl quinazolinone derivatives *Cell Pathways Inc.* **2003**, *US6562830 (B1)*.

20. Wensbo, D.; Louise, E.; Methvin, I.; McLeod, D. A.; Slassi, A.; Xin, T.; Stormann, T. M. *Astrazeneca AB., NPS Pharma Inc.* **2005**, *WO2005066155 (A1)*.

21. Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S.

Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012**, *149*, 214-231.

22. Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Crystallogr.* **1993**, *26*, 795-800.

23. Bailey, S. The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallogr. D* **1994**, *50*, 760-763.

24. Mccoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

25. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **2002**, *58*, 1948-1954.

26. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y. C.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E. Selective inhibition of BET bromodomains. *Nature* **2010**, *468*, 1067-1073.

27. Saitou, N.; Nei, M. The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4*, 406-425.

28. Zuckerkandl, E.; Pauling, L. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* Academic Press: New York, 1965; pp 97-166.

29. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725-2729.

30. Ferguson, F. M.; Fedorov, O.; Chaikuad, A.; Philpott, M.; Muniz, J. R. C.; Felletar, I.; von Delft, F.; Heightman, T.; Knapp, S.; Abell, C.; Ciulli, A. Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain. *J. Med. Chem.* **2013**, *56*, 10183-10187.

31. Velazquez-Campoy, A.; Freire, E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* **2006**, *1*, 186-191.

32. Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329-336.

33. Quinn, E.; Wodicka, L.; Ciceri, P.; Pallares, G.; Pickle, E.; Torrey, A.; Floyd, M.; Hunt, J.; Treiber, D. Abstract 4238: BROMOscan - a high throughput, quantitative ligand binding platform identifies best-in-class bromodomain inhibitors from a screen of mature compounds targeting other protein classes. *Cancer Research* **2013**, *73*, 4238.

34. Fedorov, O.; Lingard, H.; Wells, C.; Monteiro, O. P.; Picaud, S.; Keates, T.; Yapp, C.; Philpott, M.; Martin, S. J.; Felletar, I.; Marsden, B. D.; Filippakopoulos, P.; Muller, S.; Knapp, S.; Brennan, P. E. [1,2,4]Triazolo[4,3-a]phthalazines: Inhibitors of Diverse Bromodomains. *J. Med. Chem.* **2014**, *57*, 462-476.

35. Rooney, T. P. C.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin, S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M. H.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Muller, S.; Knapp, S.; Brennan, P. E.; Conway, S. J. A Series of Potent CREBBP Bromodomain Ligands Reveals an Induced-Fit Pocket Stabilized by a Cation-pi Interaction. *Angew. Chem. Int. Ed.* **2014**, *53*, 6126-6130.

36. Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P. C.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Muller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc.* **2014**, *136*, 9308-9319.

37. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430-431.

38. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **2007**, *6*, 881-890.

Chapter 6

Conclusions

MD simulations have been employed throughout this thesis over a wide range of applications. It includes ligand dissociation, protein dynamics, peptide-protein interactions and drug design, covering both theoretical and practical purposes. Depending on the research interests, various analytical tools are applied to extract the information from the simulation data, which provides us with the direct insight of the target systems. Altogether, we are presented with a vivid picture of complex processes between biological macromolecules.

The dissociation of DMSO from FKBP seemed to be so inconspicuous that we might easily underestimate its effect. Indeed, it can only cause a small perturbation to the receptor as well as the surrounding solution. However, the difference between the bound state and unbound state are statistically significant. And the time scale of the global rearrangements is two magnitude slower than the local unbinding event. This slow response implies the conformational memory of proteins, but its impact on the binding kinetics still needs to be verified. Finally, we propose that the movement of the first-layer water molecules should be coupled with protein and it should be a general rule in biophysics.

The interaction between a PDZ domain and a typical C-terminal peptide is studied using microsecond-long MD simulations. Based on the thorough and detailed analysis, we learn the kinetics profile and capture the rate-limiting step. Our conclusion is aligned with experimental data, which shows that the “encounter complex” is essential for the stereospecific interaction. In the future, we can also dig into the alternative binding mode as mentioned in the introduction. Since the electrostatic energy can no more assist the association process, enhanced sampling method like PIGS might be the essential tool.

In the project of virtual screening of CREBBP bromodomain inhibitors, MD simulations offer us precious guidance in many aspects. They help us in discriminating real binders, validating predicted binding poses and assessing the stability of critical factors like hydrogen bonds. The *in silico* computations largely reduce the cost that we would have spent for *in vitro* or *in vivo* tests. With the rapid development of force fields and algorithms, the day might come soon, in which we can use MD simulation to do high-throughput virtual screening efficiently and predict the binding free energy accurately.

In summary, MD simulations are once again proven to be a masterpiece in the toolbox for biophysical researches, by which we can explore the dynamic features and disclose the molecular mechanisms that are not accessible otherwise.

Acknowledgements

First of all, I want to thank Amedeo Caflisch for giving me this opportunity to do research in his group. I really appreciate all the support and precious advices that he has given for my work. It is an amazing experience to study in Zurich and none of these could come true without him.

I feel really grateful to the Caflisch group. Everyone is so kind and willing to help each other. Especially, I want to thank Danzhi and Andreas for all the guidance in my research. Thank Jing, Emilie and Dimitrios for the experimental part of the bromodomain project. Thank Marco for working with me on the MSM project even though we did not succeed in the end. I always enjoy the time spending with my colleagues and I know that I cannot make it through without them.

Of course, I also want to thank Prof. Peter Hamm for collaborating with me on the FKBP project. I have learnt a lot from him.

Finally, I want to thank my family and friends for always being supportive. It makes me feel much safe to explore in this unknown world while knowing that they are always there for me.

List of publications

- Bloechliger, N., **Xu, M.** and Caflisch A. Peptide Binding to a PDZ Domain by Electrostatic Steering via Non-Native Salt Bridges. *Biophys J.* 2015 May 5; 108(9): 2362-7

- Unzue, A., **Xu, M.**, Dong, J., Wiedmer, L., Spiliotopoulos, D., Caflisch, A., Nevado, C. Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *Journal of Medicinal Chemistry* 2015

- **Xu**, **M.,** Unzue, M., Dong, J., Spiliotopoulos, D., Nevado, D. and Caflisch A. Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics. *Journal of Medicinal Chemistry* 2015

- **Xu, M.,** Hamm, P. and Caflisch A. Protein Structural Memory Influences Ligand Binding Mode(s) and Unbinding Rates. *Journal of Chemical Theory and Computation* 2016

Curriculum Vitae

Min Xu

Born April 5, 1987 in Shanghai, China

Education

October 2012 – December 2015

PhD student in the group of Prof. Dr. Amedeo Caflisch, Department of Biochemistry, University of Zurich

September 2009 – June 2012

Master degree in Genetics, School of Life Sciences, Fudan University

September 2005 – June 2009

Bachelor in Biotechnology, School of Life Sciences, Fudan University